Detection of mass lesions in mammograms by using multiple views

Saskia van Engeland

Print: Universiteitsdrukkerij Technische Universiteit Eindhoven
Cover: Paul Verspaget Grafische Vormgeving
Penguin: the official Linux logo by Larry Ewing

The printing of this thesis was kindly supported by the Dutch Cancer Society.

ISBN 90-9020744-9

Detection of mass lesions in mammograms by using multiple views

een wetenschappelijke proeve op het gebied van de Medische Wetenschappen

Proefschrift

ter verkrijging van de graad van doctor

aan de Radboud Universiteit Nijmegen

op gezag van de Rector Magnificus prof. dr. C.W.P.M. Blom

volgens besluit van het College van Decanen

in het openbaar te verdedigen op maandag 18 september 2006

des namiddags om 1.30 uur precies

door

Saskia van Engeland

geboren op 14 november 1976

te Eindhoven

Promotor:	Prof. dr. C.C.A.M. Gielen
Copromotor:	Dr. ir. N. Karssemeijer
Manuscriptcommissie:	Prof. dr. ir. B.M. ter Haar Romeny (TUE)
	Prof. dr. J.G. Blickman
	Dr. T.M. Heskes

This research was made possible by Grant KUN 2001-2380 from the Dutch Cancer Society and sponsored by fundraising activities of cycling club ”Bergh in het zadel”.

Contents

Detection of mass lesions in mammograms by using multiple views

Chapter 1

Introduction

1.1 Breast cancer

Breast cancer is the most common type of cancer in women world-wide. In the western world ten percent of all women are confronted with breast cancer in their lives. Breast cancer is the result of uncontrolled growth of breast cells. The female breast is mainly made up of lobules (milk-producing glands), ducts (milk passages that connect the lobules to the nipple), fatty and connective tissue surrounding the ducts and lobules, blood vessels, and lymphatic vessels (see Fig. 1.1). Most breast cancers have their origin in the cells of the ducts, some in the cells of the lobules. The early stage of ductal cancer is referred to as in-situ, meaning that the cancer remains confined to the ducts (ductal carcinoma in-situ). When it has invaded the surrounding fatty tissue and possibly has also spread to other organs, it is referred to as invasive. Breast cancer is most treatable when it is detected in its early stages. Therefore, many nations have started breast cancer screening programs.

1.2 Breast cancer screening

The organization of a screening program is such that asymptomatic women are invited for X-ray examination of both breasts (mammogram) on a regular basis. The age range of women, and the frequency of invitation varies between different programs. After the menopause, the breast tissue gradually becomes less dense, making it easier to detect cancers. X-ray images of the breast can be taken from different angles. The most common projections are the mediolateral oblique (MLO) view and the craniocaudal (CC) view (see Fig. 1.2). The MLO view is taken from an angled view, and shows part of the pectoral muscle. The images are interpreted by one or two radiologists. Breast cancer

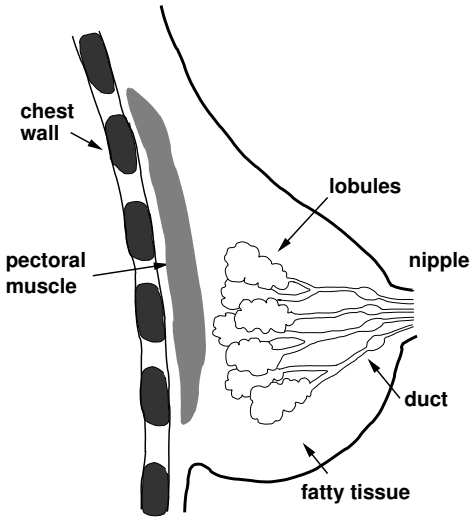
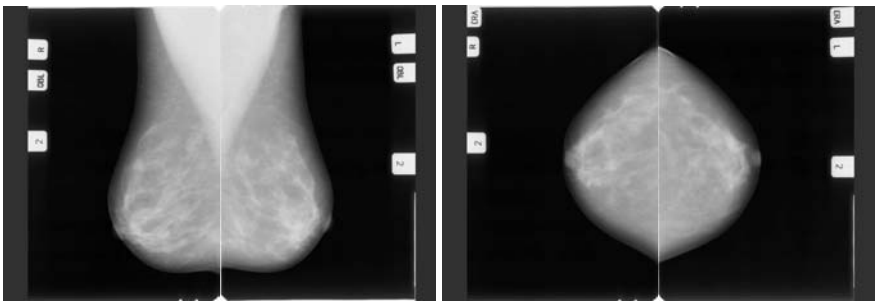
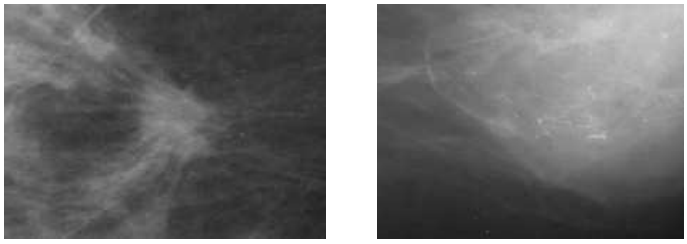


Figure 1.1: *Side view of the female breast showing the anatomy.*



MLO right MLO left CC right CC left

Figure 1.2: *A mammogram showing the two most common projections, mediolateral oblique (MLO) and craniocaudal (CC). In the MLO views part of the pectoral muscle is visible in the upper part of the images.*



mass calcifications

Figure 1.3: *Breast cancer will manifest itself in the form of a mass and/or calcifications in the mammogram. A* **mass** *is a group of cells clustered together more densely than the surrounding tissue. A mass can be caused by benign breast conditions or by breast cancer. The size, shape, and margins of the mass help the radiologist in evaluating the likelihood of cancer.* **Calcifications** *are calcium deposits that appear as white spots on the mammogram. They can occur as relatively large calcium deposits (macrocalcifications) or tiny specks of calcium (microcalcifications). Macrocalcifications are associated with benign conditions, caused by things like aging of the breast arteries, old injuries, or inflammations. For microcalcifications, the shape and layout help the radiologist judge how likely it is that cancer is present.*

will manifest itself in the form of a mass and/or calcifications in the mammogram (see Fig. 1.3). When there are abnormal findings the women is recalled for further examination, including further imaging by for instance ultrasound or biopsy.

In the Netherlands, breast cancer screening started in 1989, and in 1997 it reached full coverage of the country. Women between the ages of 50 and 75 are invited biannualy to have a mammogram taken. Participation is high, with an overall attendance rate of about 80 %. For the initial screening both a MLO and CC views are taken. For subsequent rounds the radiographer decides whether it is necessary to acquire a CC view, for instance when there are considerable changes compared to the previous (or prior) mammogram. Acquisition of a CC view occurs in about 30 % of subsequent screening rounds. The mammograms are interpreted by two radiologist independently. If one radiologist recommends recall and the other does not, then the final decision for recall is reached by discussion (consensus double reading). The recall rate of 1.3 % (in 2001) is the lowest world-wide, compared to 2-5 % for other European countries (Otten *et al.* 2005), and much higher in the US. For that same year the overall detection rate was 5 ‰.

The introduction of the breast cancer screening program has contributed to the reduction of the breast cancer mortality rate in the Netherlands. In 1986-1988, before the Dutch screening program began, the breast cancer mortality rate in Dutch women aged 55-74 years was 105.2 per 100 000; in 2001, after the screening program was in place, the mortality rate was 85.3 per 100 000 (Otto *et al.* 2003).

1.3 Computer-aided Detection

The reduction of breast cancer mortality due to screening might become even larger, since studies have shown that radiologists fail to recall a significant number of cases with breast cancer (for example Blanks *et al.* (1998)). The causes of these false negative screening mammograms are not clear. It is often suggested that abnormalities that are clearly visible in retrospect must have been overlooked. However, a cancer might also be missed because signs were misinterpreted. To increase the detection rate Computer-aided Detection (CAD) systems are being developed. These systems use pattern recognition techniques to identify features in an image that are characteristic for breast lesions. CAD systems are not intended to replace the radiologist, but rather to aid the radiologist during image review by prompting suspicious regions. Most CAD systems aim at avoiding detection errors. However, there is also evidence that CAD systems can increase the radiologists' performance by helping with the interpretation of detected lesions (Chan *et al.* 1999; Jiang *et al.* 1999; Veldkamp *et al.* 2000; Karssemeijer *et al.* 2003).

1.3.1 Benefit of CAD

A number of studies have investigated the benefit of using CAD systems. These studies can be divided into three groups. The first group of studies are the *retrospective* studies. Here the researchers test the CAD algorithms on previous mammograms of breast cancers detected by mammography. An example of such a study is that of Burhenne *et al.* (2000). In this study a panel of radiologists reviewed mammograms from 427 cases taken prior to the screening mammograms that led to the detection of breast cancer. At retrospective review, they found that 67 % of the screen detected breast cancers were visible on the prior mammogram, and 27 % (n=115) were deemed actionable. Of these 115 cases, the CAD system correctly marked 77 %. These results suggest that CAD prompting could have potentially reduced the false negative rate by 77 %. The second group of studies are the *prospective sequential* studies. Here the mammograms are first interpreted by a group of radiologists without the assistance of CAD, followed immediately by a re-evaluation of areas marked by the CAD system. In such a study, conducted by Freer & Ulissey (2001), 12,860 mammograms were read by radiologists with and without the assistance of CAD. When comparing the radiologists' performance without

CAD with that when CAD was used, they observed both a 19.5 % increase in the number of cancers detected and an improvement in the proportion of malignancies detected at an earlier stage. The increase in recall rate from 6.5 % to 7.7 %, was considered to be quite acceptable by the authors. The last group of studies are the *prospective comparison* studies. In these studies the changes in recall and cancer detection rates are assessed after the introduction of a CAD system. An example of such a study is that of Gur *et al.* (2004). In their study they compared the recall and breast cancer detection rates for 24 radiologists who interpreted 115,571 screening mammograms with (n = 59,139) or without (n = 56,432) the use of a CAD system. They found no statistically significant changes in recall and breast cancer detection rates. There was a lot of debate about the meaning of these results. Feig *et al.* (2004) made some important points regarding the study by Gur *et al.* (2004). One of these points was that the large confidence intervals allow for a wide range of detection rate changes. The source of this large variation is the result of variability among readers, and the fact that the cases read with and without CAD were different. Another important point made was that the percentage of women who were screened for the first time decreased from approximately 40% to 30%. A population undergoing periodic mammographic screening typically has lower cancer detection rates on second and third screens than on the first screen. As a consequence, this may have obscured any evidence of benefit from CAD.

1.3.2 Multiple view CAD techniques

In general, the sensitivity of CAD techniques for mass detection is high, however the specificity is only modest. To reduce the number of false positive detections, there is a lot of interest in developing CAD techniques that use image context, asymmetry, and multiple view information. Most current CAD schemes use information from only one view at the time. For a case based evaluation of CAD performance the results of all views are independently combined afterwards. Radiologists on the other hand use the information from all available views during interpretation. They are trained to use comparisons of the left and right breast to identify suspicious asymmetric densities. The comparison of temporal mammogram pairs is important to see whether there are new or growing densities in the mammogram. It is also known that screening with two mammographic views, MLO and CC, improves the detection accuracy of abnormalities in the breast, which can be explained by the fact that two projections allow better estimation of conspicuity of lesions and may reveal lesions hidden by glandular tissue in one of the projections.

In literature a number studies have been described that investigate the use of multiple view information for CAD techniques, including the search for asymmetry, temporal comparison and combination of MLO and CC information. Asymmetry detection has been performed by creating a mapping between the left and right breast using a set of control points on the skin line (Karssemeijer & te Brake 1998; Lau &

Bischof 1991). A small benefit was found when using asymmetry in the detection scheme. However, no studies have been published since then that confirm the usefulness of asymmetry for automated detection of breast lesions. With respect to the use of temporal information, Sanjay-Gopal *et al.* (1999) developed an automated regional registration technique to identify corresponding lesions in temporal pairs of mammograms, and Hadjiiski *et al.* (2001b) developed interval change analysis of masses to improve their classification accuracy between benign and malignant. For the use of MLO and CC information a number of studies have been presented on methods to establish correspondence between detected objects in both views (Highnam *et al.* 1998; Good *et al.* 1999). To our knowledge, the only study presented on the use of MLO and CC information to improve detection results is that of Paquerault *et al.* (2002). They developed a two view matching method which results in a correspondence score for each possible mass pair. By combining this correspondence score with their single view detection score, their detection results improved significantly.

1.4 Breast density

There are various risk factors connected with breast cancer. It is well known that increasing age, hormonal factors, especially a long menstrual life, and no or late childbearing, have an important influence. The use of oral contraceptives at a young age and the use of hormone replacement therapy are also known to increase the risk of developing breast cancer. The occurrence of breast cancer in first degree relatives also increases the risk. In some families the patterns of breast cancer incidence seems to be consistent with known patterns of genetic inheritance. In 1994, researchers discovered that women who carry mutations of BRCA1 (breast cancer gene 1) or BRCA2 (breast cancer gene 2) are at higher risk of developing both breast and ovarian cancer than women who do not have these genetic mutations (Futreal *et al.* 1994). Research is still continuing to gain more insight in risk factors.

For many years, research has focused on breast tissue density as an important risk factor for breast cancer development (Byrne *et al.* 1995). Some studies even showed that the risk of breast cancer for women with increased breast density is four to six times that for women with less dense breasts (Harvey & Bovbjerg 2004). Breast tissue density plays an important role in epidemiological studies, and might be used to identify a high risk population. This high risk population might then benefit from more frequent screening and/or additional imaging with other modalities. In addition, including breast density as a feature in a CAD system might be used to increase performance in CAD systems. In current practice, breast density is determined qualitatively by the radiologist. The use of breast density in CAD systems requires, however, a less subjective and more quantitative measure of breast density that can be obtained automatically.

1.5 This thesis

The topic of this thesis is the development of computer-aided detection and image processing techniques that improve the detection of masses in mammograms by using multiple view information. Chapters 2 and 3 address the comparison of temporal mammogram pairs. Chapters 4 to 6 address the improvement of CAD results by using two views, MLO and CC, of the same breast.

To be able to make temporal comparisons it is important that previous and current mammograms are similar. When reading a mammogram, radiologists use the mammogram from the previous screening round to search for growing or new densities. To facilitate this, correct positioning of the breast during the examination is very important. Inadequate positioning is, however, a frequently encountered problem. To some extent suboptimal positioning can be compensated for by registration. In Chapter 2 we tested four methods for temporal mammogram registration. Two of the methods were relatively simple alignment procedures based on nipple location and center of mass of the breast area. The third method used mutual information, which is computed from the joint probability distribution of the images' intensities. And the fourth method used a warping approach, based on a set of automatically determined control points located on the breast contour.

In mammography, it is generally believed that growth of lesions is an important clue to detect lesions and to discriminate between benign and malignant. The introduction of digital mammography over the last few years has given rise to new opportunities. Digital mammography also opens new ways to design display strategies for screening, which might also lead to improved perception of lesion growth. In Chapter 3 we investigated two ways of presenting prior and current mammograms on a mammography workstation: next to each other and alternating at the same display (toggle). Using an observer experiment, we determined threshold detection levels of lesion growth under both conditions.

During the development of multiple view techniques we have noticed that they are hampered by the phenomena that mass lesions are sometimes detected by multiple regions of our CAD scheme. These multiple candidate regions on the same lesion complicates the process of automatically linking regions in multiple views. Chapter 4 describes a technique we developed to regroup initial CAD regions to facilitate the final classification of suspicious regions. This method aims at recognizing lesions that are hit by multiple regions and performing a re-segmentation and re-evaluation as a single region.

It is common for CAD algorithms to combine detection results in separate views independently. However, most development of CAD systems has been based on the analysis of single views. Before multiple view information can be used, correspondence has to be established between views. In Chapter 5 we present a method to link potentially suspicious areas of our CAD scheme in MLO and CC views. Using this linking method, Chapter 6 focuses on the improvement of the mass detection results using two view

information.

A number of studies investigated the use of image context for CAD systems, including asymmetry and multiple view information. Concluding from these studies, however, the benefit of context and multiple view information for CAD systems is not always clear. In Chapter 7 we present a study in which we made use of human observers to investigate the potential benefit of using context information for CAD systems. The observers were presented with a number of cases in three different ways, showing only the region of interest, showing the whole image, and showing all available images of the screening round. They were asked to rate each presented view by giving it a malignancy score.

As mentioned in the previous section, breast tissue density has been identified as an important risk factor for breast cancer development. In literature a number of studies have been presented that determine the size of the dense tissue projection on a mammogram. As a quantitative measure of breast density the projected dense tissue area is, however, not ideal, as it is not invariant to compression and projection angle. The last chapter of this thesis presents a method for accurate quantitative estimation of the dense tissue volume from mammograms obtained with full field digital mammography. We validated this method by comparing with the dense tissue volumes determined from Magnetic Resonance Imaging data.

Chapter 2

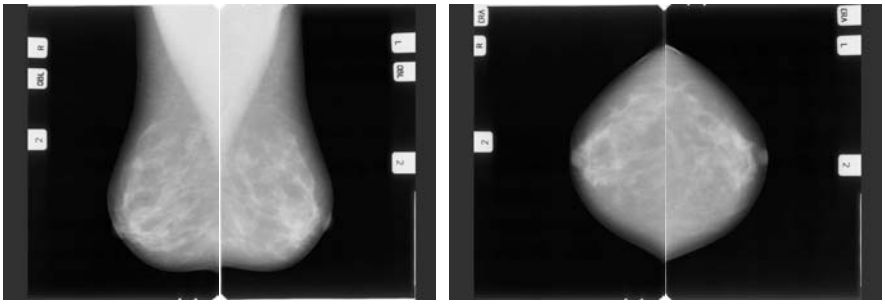
A comparison of methods for mammogram registration [1]

2.1 Introduction

Breast cancer is the most common type of cancer in women world-wide. Ten percent of all women are confronted with breast cancer in their lives. When it is detected in its early stages, breast cancer is most treatable. Therefore, many nations have started mammography screening programs to detect breast cancer as early as possible. The organization of a screening program is such that asymptomatic women are invited for X-ray examination of both breasts on a regular basis. The age range of women, and the frequency of invitation varies between different programs. After the menopause, the breast tissue gradually becomes less dense, making it easier to detect cancers. X-ray images of the breast can be taken from different angles. The most common projections are the mediolateral oblique (MLO) view and the craniocaudal (CC) view (see Fig. 2.1). The MLO view is taken from an angled view, and shows part of the pectoral muscle. The images are interpreted by one or two radiologists. Breast cancer will manifest itself in the form of a mass or calcifications in the mammogram. When there are abnormal findings the women is recalled for further examination. During interpretation information from all available views is used. The mammograms from previous examinations are used to recognize changes in the breast, and the mammogram is searched for asymmetries.

Since radiologists make comparisons between different views, both bilateral and temporal, positioning of mammograms is important to aid the radiologist during mammogram interpretation. To facilitate scanning for asymmetries, positioning should be such that the nipples of the left and right breast are projected on a horizontal line. Further-

[1] The content of this chapter has been published previously in van Engeland *et al.* (2003a).



MLO right MLO left CC right CC left

Figure 2.1: *A mammogram showing the two most common projections, mediolateral oblique (MLO) and craniocaudal (CC). In the MLO views part of the pectoral muscle is visible in the upper part of the images. When breast positioning is correct the nipples are horizontally aligned and the pectoral muscle is equally visible in the left and right breast. Left and right images are hung next to each other to create a symmetric view.*

more, the pectoral muscle and glandular tissue should be projected as similar as possible for left and right breasts. Also for temporal comparisons it is important that previous and current mammograms are similar. For example, it has been shown that alternating between previous and current views at the same display improves the perception of lesion growth (van Engeland *et al.* 2003b) (see Chapter 3). Proper positioning of the breast requires highly skilled radiographers. Inadequate positioning is the most frequent problem encountered when reading mammograms (Dronkers *et al.* 2002). A mammogram which is properly positioned is displayed in Fig. 2.1. Introduction of digital mammography and soft-copy reading of mammograms enables the use of display optimization techniques. To some extent suboptimal positioning can be compensated for by registration.

Registration of mammograms is not only important in displaying mammograms but it can also be used to develop Computer-aided Detection (CAD) methods that use information from previous mammograms. Most current CAD techniques for mammography are based on the analysis of single views and detection. However, interest in CAD for classification into benign and malignant has grown, and for this purpose temporal analysis is usually required. In addition to improving classification, also detection might be improved by using information from previous mammograms.

The mammographic appearance of breast tissue in previous and current mammograms of the same patient may vary considerably, because of differences in breast compression and positioning, differences in imaging techniques, and changes in the breast itself. As mentioned earlier, after the menopause the dense glandular tissue starts disappearing. Together with the fact that there are no clear landmarks in a mammogram, except for the nipple, this makes mammogram registration a challenging task.

In literature some approaches have been described for mammogram registration and techniques to find corresponding lesions in pairs of mammograms. Most work has been done on the registration of temporal pairs of mammograms. Sallam & Bowyer (1996) used a two-dimensional warping technique to find correspondence between all pixels within the breast region in the current image to pixels within the breast region in the reference image. Wirth *et al.* (2002) combined the use of a similarity-measure and a point-based spatial transformation for the registration of temporal mammogram pairs. In contrast to these studies where the whole breast area is registered, Hadjiiski *et al.* (2001a) developed an automated regional registration technique to identify corresponding lesions in temporal pairs of mammograms. A bilateral registration technique was developed by Yin *et al.* (1994). This technique was based on determining correspondence between control points on the skin line. Corresponding points of the right breast were registered with those on the left breast using a least-squares method, allowing translation and rotation. Validation of the registration methods mentioned above was only presented by Hadjiiski.

We implemented and validated four methods for temporal mammogram registration. The first two methods were relatively simple alignment procedures based on nipple location and center of mass of the breast area. For the first method the nipple was detected automatically and the nipple in the previous and current view were positioned at the same location by translation of the previous view. For the second method the mammograms were segmented into breast, pectoral and background area, and the center of mass of the breast area in the previous and current view were positioned at the same location by translation of the previous view. The third method used mutual information, which is computed from the joint probability distribution of the images' intensities. By maximizing mutual information, two images are registered. In contrast to the first two techniques, also rotation, scaling and shearing could be applied to register previous and current views. The fourth technique was a warping approach. This technique used a set of automatically determined control points located on the breast contour and the pectoral muscle. The control points in both views were mapped onto each other, and the warped image was obtained by interpolating between the control points using a thin-plate spline surface method.

In many applications validation of registration methods is problematic, as a ground truth is rarely available. In this study we take a task-oriented validation approach. As the main purpose of mammogram registration is lesion examination, we used the adequacy of the alignment of lesions in temporal pairs as our validation measure. To this end, we used a set of 150 previous and current mammographic views from the Dutch breast cancer screening. Densities and calcifications visible in both views were annotated. The distance between the centers of mass of these annotations after registration was used as a criterion to compare the methods.

2.2 Methods

2.2.1 Data set

The data set we used consisted of 150 temporal pairs from the Dutch breast cancer screening program. The mammograms were digitized with a resolution of 50 μm and averaged down to 200 μm (12 bits/pixel). Annotations of landmarks, either mass-like lesions, in the rest of the paper referred to as densities, or large calcifications visible in both previous and current view, were made. The densities we selected were mostly benign, and in the range of 0.5 - 1.5 cm in diameter. The data set contained: 60 densities and 90 calcifications visible in 30 CC views, and 120 MLO views. The average time interval between the previous and current mammographic views was 23 months (range 6-27 months).

2.2.2 Nipple alignment

The first method we tested was the alignment of previous and current views by registering the nipples, by translating the images in horizontal and vertical direction. This approach is similar to the approach used by many radiologists, who use the distance to the nipple to find corresponding locations of lesions in two mammographic views. Radiologists we interviewed had the experience that this distance remains fairly constant between views.

We used an automated nipple detection method. The first step in this method was the segmentation of the mammogram into background and breast tissue. Our breast segmentation algorithm, which is similar to the method described by Highnam & Brady (1999), is based on the idea that the background is very flat and can be identified by low gradient values. We use a fixed threshold on the gradient magnitude to determine pixels in the background. In addition, a threshold on the image intensity values is used to exclude brighter pixels from the background. This threshold is determined by histogram analysis. Morphological operations are applied to merge regions in the breast and to remove labels. By assuming that the location of the nipple is somewhere on the skin contour, the problem of locating the nipple was reduced to a one-dimensional problem. For every point on the contour, features that describe intensity gradients, shape of the skin contour,

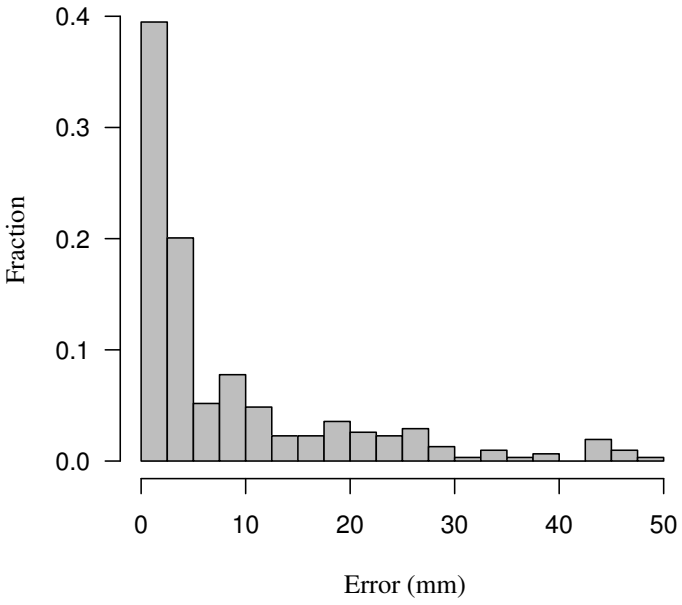


Figure 2.2: *Histogram of the distance from the actual to the estimated nipple position for the used data set.*

line patterns of the glandular tissue, and the geometry of the breast, were determined and fed into a three-layer backpropagation network. The network was trained using a set of 314 mammograms with known nipple position. The distance to the actual nipple location (tip of the nipple), which was indicated by a non radiologist working in the field of mammography, was used as the target value for the network. Figure 2.2 gives the nipple detection results on the testing set of 300 mammographic views that were used for the comparison of the registration methods.

Registration based on nipple location has some disadvantages. First, the detection of the nipple depends strongly on the correct segmentation of the mammogram. When the breast contour is not detected properly, the nipple detection method may easily give a wrong nipple location. Second, even when the breast contour is detected correctly, this registration technique will only work when the breast is positioned correctly. For a correctly positioned breast the nipple is visible in profile. In our data set, we have found that in 35 % of the images, the nipple was not clearly visible.

To study the potential of this method, we also made a 'best case' comparison using manually indicated nipple positions. Finally, to investigate the importance of correct breast positioning for this registration technique, our data set was split up into two sets, one set where the nipple was clearly visible in profile in both temporal views and one set where the nipple was not clearly visible in one or both temporal views. For this comparison we used the network determined nipple positions.

2.2.3 Center of mass alignment

Mammographic views can also be registered by aligning the centers of mass (CM) of the breast area. However, analogous to the previous method, the performance of this method will depend strongly on segmentation of the mammogram. Especially with high contrast films used nowadays, it is often very difficult to find the breast contour. Another problem with center of mass alignment in the MLO views is the varying proportion of the pectoral muscle that is visible. When positioning is not adequate, this may give a large variation in the vertical coordinate. Therefore, we expect that this method can be improved by excluding the pectoral muscle in the center of mass calculation. We implemented registration with and without the pectoral muscle, where the edge of the latter was detected automatically.

The pectoral muscle segmentation technique is described in detail in Karssemeijer (1998). It selects a region of interest in the image where the pectoral muscle is likely to be located. By applying the Hough transform on this region a parameter space, which represents straight lines through the region, is obtained. Finally, the highest peak is selected, which represents the straight line estimate of the pectoral muscle boundary. To improve the segmentation, pixels with a stronger gradient have a larger increment in the Hough accumulator, and a correction is made to reduce the bias towards longer pectoral muscle boundaries. Furthermore, if there are more candidate peaks in Hough space, the peak with the largest corresponding pectoral muscle area is selected.

2.2.4 Mutual information

The use of mutual information (MI) for medical image registration applications was independently introduced in 1995 by both Viola & Wells (1995) and Collignon *et al.* (1995). It is a popular registration technique also used for multi-modality image registration. Since the appearance of previous and current views can be considerably different, mutual information might be a good technique for temporal mammogram pair registration. It is expected to be less dependent on positioning changes than the previous alignment techniques, as it uses the internal structure of the breast.

Mutual information gives a measure of the strength of the dependence between variables (ie. pixel values occurring in two images). For two images A and B it is defined as:

$$I(A,B) = H(A) + H(B) - H(A,B), \tag{2.1}$$

where $H(A)$ and $H(B)$ being the entropies of A and B, respectively and $H(A,B)$ their joint entropy. The entropy of image A is defined as:

$$H(A) = -\sum_{a} p_A(a) log p_A(a), \tag{2.2}$$

where $p_A(a)$ is the pixel value distribution of image A. In other words, $I(A,B)$ is the amount of information that image B contains about image A. Therefore, maximization of Eq. 2.1 by transformation of one of the images is equivalent to registering the images.

The current mammographic view is considered as the $reference\ image$, and transformations are applied to the previous view or $floating\ image$. This naming convention will be used throughout the paper. Possible transformations were translation, rotation with respect to image center, scaling and vertical shearing. This in contrast to the previous two methods, which only used translation. Scaling was included in the transformation to correct for slight size changes due to varying compression. A disadvantage of allowing scaling might be that the registration method can use scaling to compensate for physical changes, like the reduction of glandular tissue after menopause. This might make interval change analysis unreliable. This is a major drawback, as estimation of lesion growth is very important for classification of benign and malignant lesions. To study potential problems with scaling, we tested MI registration both with and without scaling. We could use a much larger data set of 1000 pairs of temporal mammographic views for this experiment, since no lesion annotations were required. We assumed that by using a large data set compression differences between subsequent views would average out. The average interval between subsequent mammograms in this series was two years. Finally, we also determined the size of the segmented breast area for all temporal pairs and compared these results with the scaling factors found by MI registration.

Mutual information uses internal structures in the breast and assumes that the co-occurrence of the most probable values in two images is maximized at registration. This might be a problem when large changes occur, for instance due to an aggressive invasive cancer. In our study database we excluded such cases, but we investigated this issue in a separate series of 20 cases with strong interval change (average lesion diameter 2.5 cm). A mask was used to exclude the regions where the lesions developed for MI value determination. For this set of image pairs no landmarks were available to determine the accuracy of MI registration. Therefore we could only compare transformation values with and without the excluding mask. One might argue that similarity of transformation values with and without exclusion mask would not imply that the registration is correct. Therefore, we also used an exclusion mask (diameter 2 cm) on the 150 image pairs for which landmarks were available, and compared these results with the registration results obtained without masking part of the breast.

Implementation To be able to exclude background pixels for MI value determination, the $reference$ and $floating\ image$ were segmented into background and breast tissue. Both images were linearly scaled to 128 pixel values. For less than 64 pixel levels bins a slight decrease in performance was found. Bi-linear interpolation was used to determine the image intensity at the transformed image position. Transformations were applied in the following order: shearing, rotation, scaling and translation. Only vertical shearing was applied and rotation was performed with respect to the image center. The MI value was optimized by transforming the $floating\ image$ using Powell optimization (Press *et al.* 1992).

2.2.5 Warping

The fourth method we tested was a method originally developed to define a mapping between the left and the right breast to obtain a local measure of asymmetry (Karssemeijer & te Brake 1998). This technique used a set of automatically determined control points (n_{mlo}=14 and n_{cc}=10) located on the breast contour and the pectoral muscle (Fig. 2.3). The control points were obtained after segmenting the mammograms into breast tissue, pectoral muscle and background. After mapping the points in both views onto each other, the warped image was obtained by interpolating between the control points using a thin-plate spline method (Bookstein 1989).

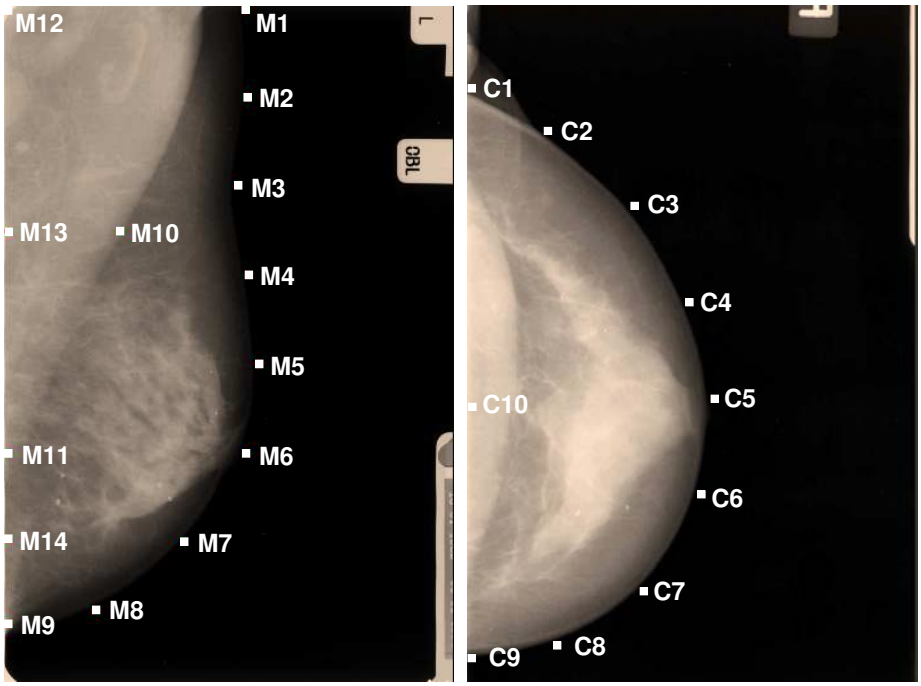


Figure 2.3: *Control points in a MLO and a CC view. On the skin line both end points (***M1***,* **M9***,* **C1** *and* **C9***) and seven equidistant points (***M2** - **M8** *and* **C2** - **C8***) were selected. For the MLO view two points on the pectoral boundary were added, respectively one in the middle (***M10***) and one at the intersection with the chest boundary (***M11***). At the chest boundary three more control points were selected, one at the top (***M12***), one (***M13***) in between* **M11** *and* **M12***, and one (***M14***) in between* **M9** *and* **M11***. For the CC view only one control point (***C10***) at the chest boundary in the middle of both end points (***C1** *and* **C9***) was added.*

2.3 Results

2.3.1 Comparing results from different techniques

Figure 2.4 and Table 2.1 show the registration results obtained with the four techniques. CM registration results were obtained after excluding the pectoral muscle, and MI results were obtained using translation and rotation. The MI results presented in the table were obtained using translation and rotation. MI outperformed the other registration techniques, and the warping technique performed worst. The main reason for this is that the warping algorithm only takes into account points on the breast contour. Registration by center of mass turned out to be better than registration by nipple location.

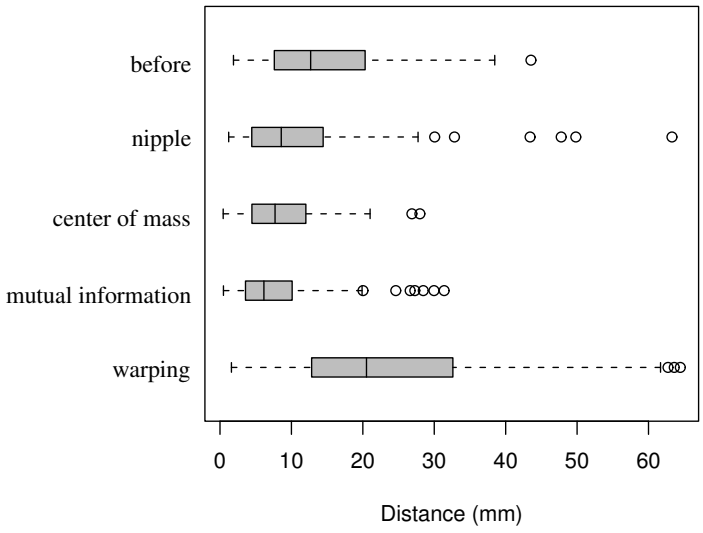


Figure 2.4: *Boxplot of distances in mm between lesion in previous and current view before and after registration using different techniques. Boxes show lower quartile, median and upper quartile values. The whiskers indicate the minimum and maximum distances. ○ represent outliers, which are data points with a distance to the upper quartile value of more than 1.5 times the interquartile range.*

In Fig. 2.5 registration based on manually indicated nipple positions is compared with registration based on network determined nipple positions. The median registration results are about equal, however the number of outliers is higher for registration based on network determined nipple positions. Registration based on nipple position was better for cases where the nipple was clearly visible in profile (Fig. 2.6), this difference was, however, not significant (Wilcoxon rank sum test, p = 0.11).

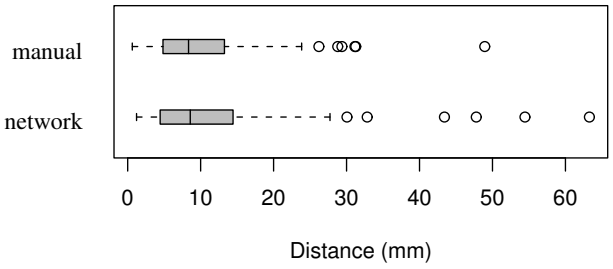


Figure 2.5: *Boxplot of distances in mm between lesion in previous and current view after registration using alignment by nipple location, where the nipple was either manually indicated or detected using the neural network.*

In Fig. 2.7 center of mass alignment results are compared with and without excluding the pectoral muscle. It can be seen that pectoral muscle segmentation is essential for good performance of this technique. In Fig. 2.8 and Table 2.2 MI registration results are presented using different transformation settings. Table 2.3 gives the transformation values used to optimize MI.

From Table 2.3 it can be seen that the average scale factor found by MI registration is not equal to one. This result could be reproduced on a larger data set of 1000 temporal image pairs, again with an average time interval of two years. For this data set we found an average scale factor of 0.97 (significantly different from 1.00, sign test, $p \ll 0.001$). We also compared the size of the segmented breast area in the previous view with the size in the current view. This comparison resulted in an average scale factor of 0.98 between previous and current views (significantly different from 1.00, sign test, $p \ll 0.001$). However, the linear relationship between both scale factors turned out to be weak (correlation coefficient 0.40).

To study the effect of large developing densities on MI registration performance, we used 20 temporal image pairs with strong interval change and a mask to exclude the regions where the lesions developed. We compared the mean and standard deviation of the

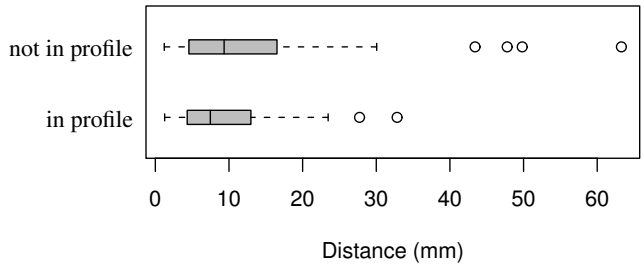


Figure 2.6: *Boxplot of distances in mm between lesion in previous and current view after registration using alignment by nipple location for the cases where the nipple was clearly visible in profile and the cases where the nipple was not visible in profile for one or both temporal views.*

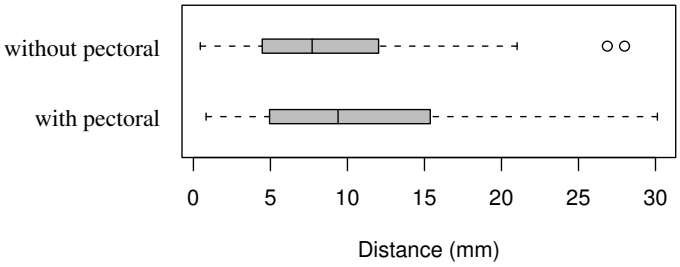


Figure 2.7: *Boxplot of distances in mm between lesion in previous and current view after registration using alignment by center of mass either with or without pectoral muscle.*

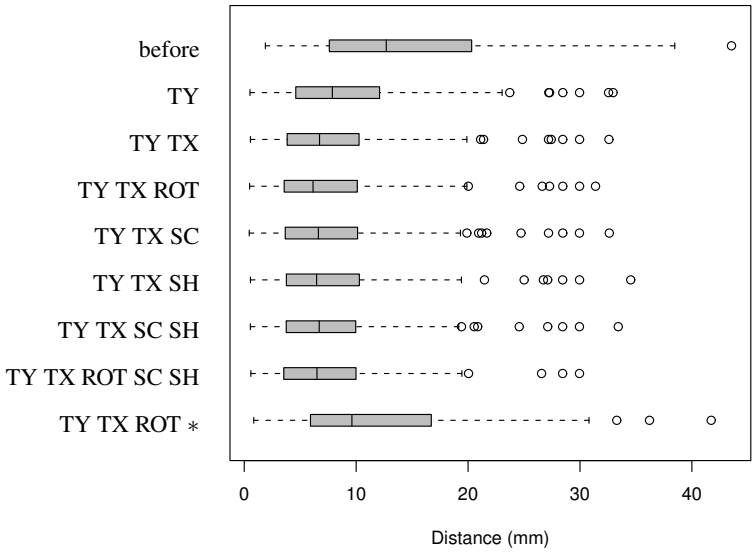


Figure 2.8: *Boxplot of distances in mm between lesion in previous and current view after registration using mutual information with different transformation settings, using a combination of horizontal (TX) and vertical translation (TY), rotation (ROT), scaling (SC) and shearing (SH). $*$ indicates that images were used without excluding the background.*

Table 2.1: *Results after registration by different techniques. Values given are minimum, median, mean and maximum distances between lesions in previous and current views and the standard deviations (SD) in mm.*

	min	median	mean	max	SD
before	1.9	12.7	14.7	43.5	8.6
nipple	1.2	8.6	11.1	63.3	9.7
center of mass	0.4	7.7	8.8	28.0	5.7
mutual information	0.5	6.1	7.9	31.4	6.2
warping	1.6	20.5	24.5	142.3	17.8

Table 2.2: *Results after MI registration with different transformation settings using a combination of horizontal (TX) and vertical translation (TY), rotation (ROT), scaling (SC) and shearing (SH). Values given are distances between lesions in previous and current views in mm. $*$ indicates that registration was done without excluding backgrounds.*

	min	median	mean	max	SD
before	1.9	12.7	14.7	43.5	8.6
TY	0.5	8.0	9.8	56.3	7.7
TY TX	0.6	6.7	8.1	32.6	6.2
TY TX ROT	0.5	6.1	7.9	31.4	6.1
TY TX SC	0.4	6.6	8.0	32.6	6.1
TY TX SH	0.6	6.5	8.1	34.5	6.2
TY TX SC SH	0.5	6.7	8.0	33.4	5.9
TY TX SC ROT SH	0.6	6.5	7.5	30.0	5.4
TY TX ROT $*$	0.9	9.6	12.0	41.7	8.0

transformation parameters with and without exclusion: TX: -1.8 $\pm$ 3.2 with exclusion, -1.8 $\pm$ -3.8 without exclusion, TY: 4.4 $\pm$ 17.0 with exclusion, 4.4 $\pm$ 17.1 without exclusion, ROT: 2.7 $\pm$ 5.0 with exclusion, 3.0 $\pm$ 5.0 without exclusion. None of these results were significantly different (paired two-sided t-test). We also compared the mean and standard deviation of the registration results with and without exclusion mask for the dataset with 150 image pairs, and found a mean distance error 7.9 $\pm$ 6.1 for both registrations. The performances of registration with and without exclusion mask were not significantly different (paired two-sided t-test). Therefore, it may be concluded that MI is not hampered by the changes between the previous and the current view.

Finally for all four techniques, registration results obtained for MLO and for CC temporal image pairs were compared. As can be seen in Fig. 2.9 the distance between annotations in temporal pairs is smaller for lesions in CC views before registration as well as after registration using the nipple, center of mass and warping. For MI registration median distance errors are equal for lesions in MLO and CC views.

2.3.2 Examples of registering temporal mammogram pairs

In this section some examples will be given of image registration results obtained by the tested methods, including cases where image registration fails, to help understand shortcomings of the techniques.

Table 2.3: *Transformation values used to optimize MI for the 150 mammographic image pairs. Transformations were applied in the following order: shearing, rotation, scaling and translation.*

	min	median	mean	max	SD
vertical shear	-0.25	0.00	0.00	0.16	0.05
rotation (deg)	-9.5	0.0	0.0	13.7	3.1
scaling	0.80	0.99	0.97	1.07	0.05
horizontal translation (mm)	-17.2	-0.1	0.3	24.0	6.0
vertical translation (mm)	-45.1	0.0	0.2	36.1	12.0

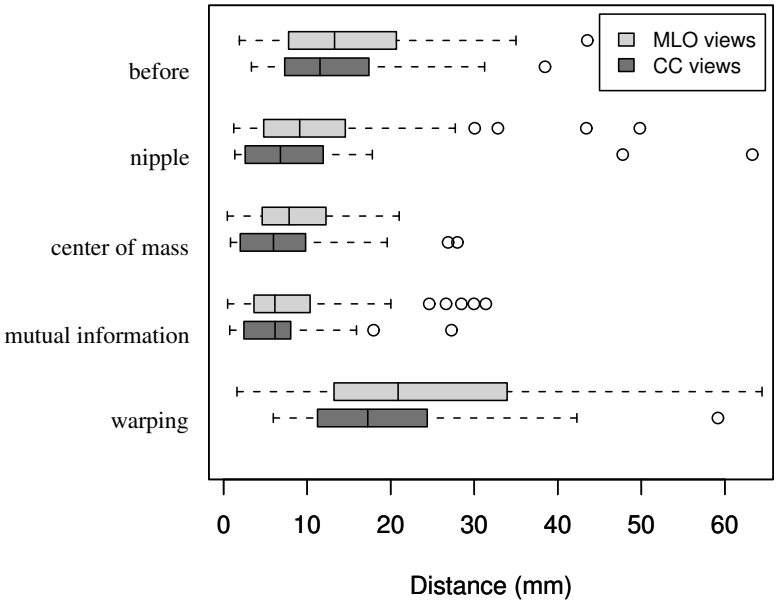


Figure 2.9: *Boxplot of distances in mm between lesion in previous and current view before and after registration using different techniques for MLO and CC image pairs.*

Nipple alignment

The assumption that the lesion to nipple distance remains constant between views holds only when the breast is positioned correctly and the nipple is visible in profile. Therefore, using registration based on nipple location is only guaranteed to work when the nipple is visible in profile (65 % of the images). The difference in registration performance for this technique between correctly and not optimally positioned cases can be seen in Fig. 2.6.

Center of mass alignment

Registration by alignment of the centers of mass of the previous and current view is a very basic approach to mammogram registration, which appeared to work reasonably well. However, we have found that segmentation of the pectoral muscle for MLO views is essential for this technique (see also Fig. 2.7). Fig. 2.11 shows the results of CM alignment with and without pectoral muscle segmentation.

Mutual information

As an example, in Fig. 2.10 the joint pixel value distributions of two temporal views before and after MI registration are presented. We found that registration by mutual information is sometimes hampered by labels or a nonuniform background in the images. Therefore, excluding the background significantly improved registration results, which can be seen in Fig. 2.8 and Table 2.2. Figure 2.12 shows an example where MI registration fails because of the presence of labels in both the previous and the current view.

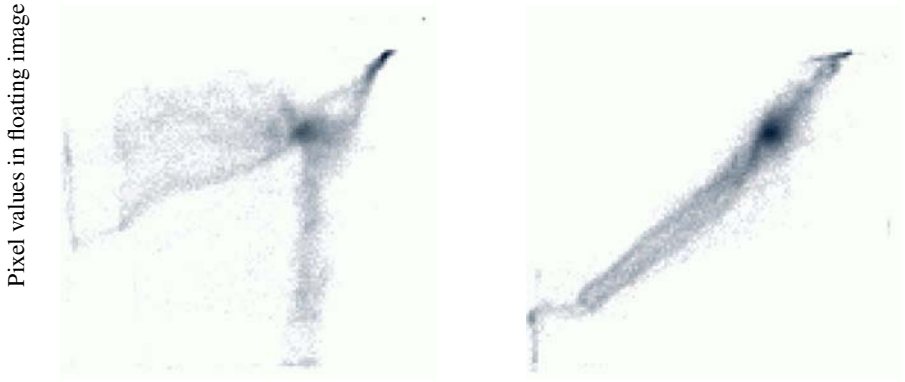
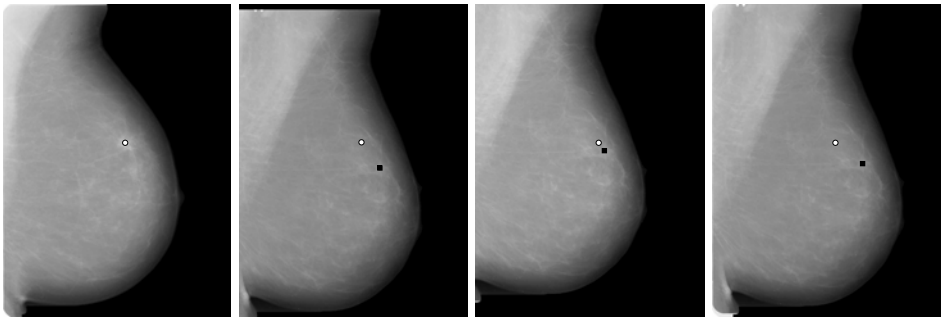


Figure 2.10: *Joint pixel value distributions of two temporal views before and after registration.*

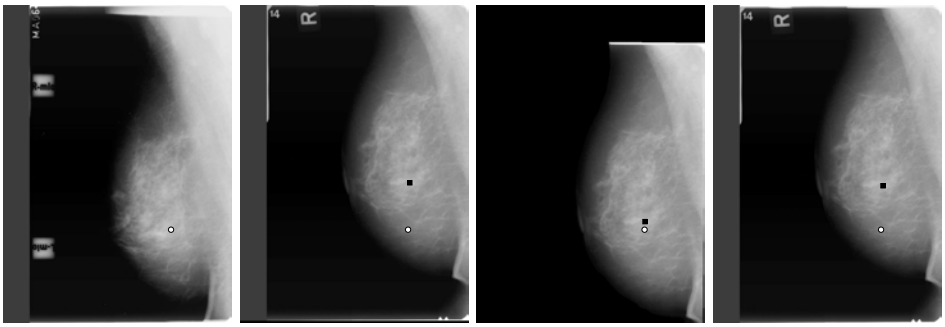


reference image | with pectoral muscle | without pectoral muscle | floating image

Figure 2.11: *Example of a previous and current image pair before and after registration using center of mass alignment with and without segmenting the pectoral muscle. This example shows the effect of excluding the pectoral muscle for CM calculation on registration. In all images the location of the lesion in the reference image is indicated (◯), also in the floating images the corresponding lesion location is marked (■).*

Warping

The warping approach we tested performed worst of all four registration methods. Because the warped image is calculated by interpolation between control points on the breast outline, strange deformations sometimes occurred inside the breast area of the warped image (see Fig. 2.13).



reference image | with background | without background | floating image

Figure 2.12: *Example of registration using MI for images with and without excluding the background. As can be seen, by excluding the background registration results improve.*

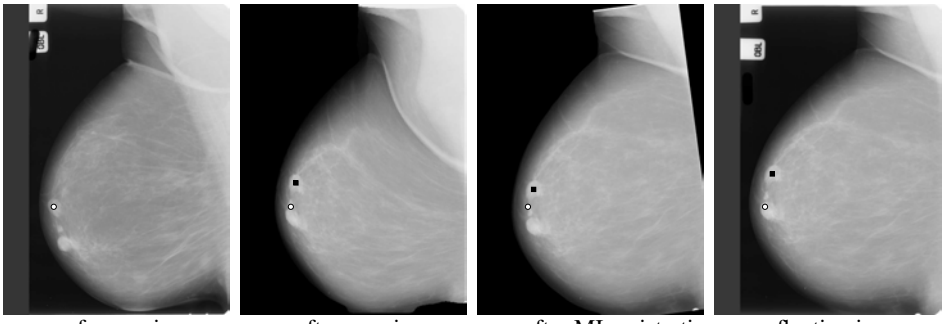


Figure 2.13: *Example of a previous and current image pair before and after registration using warping and MI registration. Strange deformations inside the breast area of the warped image can be seen.*

2.4 Discussion & conclusions

It was found that registration by mutual information outperformed other tested registration techniques. However, it appeared that mammogram segmentation is necessary to exclude the background. The benefit of mutual information over the other techniques is that it makes use of the internal structure of the breast, and not just one point or several points on the breast contour. Another advantage of this technique is that it allows rotation, scaling and shearing. This could correct for compression differences between the previous and the current mammogram and slight rotations of the breast.

The basic registration approach of aligning the centers of mass of the breast area gave reasonable results. However, we found that when using this approach pectoral muscle segmentation was essential. In practice this can be a disadvantage, because it makes the method more complicated and less robust. Our method for pectoral muscle segmentation, which uses a straight line approximation, rarely failed completely. By visual inspection, we found that for only one image in the data set used the algorithm selected a line that did not match the pectoral boundary at all. One might argue that the straight line approximation may not be accurate enough. The shape and position of the boundary might for instance be optimized by an active shape or contour model. However, some variability in the segmentation is unavoidable, since sometimes the pectoral boundary cannot be identified unambiguously even by human readers. We expect that improving pectoral muscle segmentation will not lead to a strong improvement of the registration results obtained by warping and CM alignment. Most outliers in the results obtained with these techniques were caused by large differences between subsequent temporal views.

Results obtained by registration with nipple alignment and warping turned out to be worse than the performance of the other two techniques. The main drawback of the nip-

ple registration technique is that it is very sensitive to correct positioning of the breast and the visibility of the nipple in profile. Improving the nipple detection method may remove some outliers from the registration results. However, our experiment with manually identified nipple locations showed that it will not significantly improve results. Warping might be improved by including control points inside the breast. It is very difficult to define such points. Attempts have been made by Vujovic & Brzakovic (1997) and Sallam & Bowyer (1996), who presented methods for placing control points on the intersections of elongated structures and on the boundary of dense tissue, respectively. It seemed however that the definition of these points is not accurate enough for our purpose.

We studied the scale factor found by MI registration on a large data set of 1000 temporal image pairs. As shown in Sec. 2.3.1, this scale factor is on average lower than one. In addition it can be seen in Table 2.3, that sometimes scaling factors selected are quite large, which is unacceptable. The intention of using scaling was to correct for small compression differences, however sometimes large scaling factors were selected to correct for large differences in positioning between subsequent views. Large scaling factors will make size estimations and determination of lesion growth inaccurate. If the global registration step is to be followed by a regional registration step and subsequent feature analysis for temporal CAD systems, it is safer to use MI registration without scaling. Also our results showed that there is no significant benefit in the use of scaling (Table 2.2 and Fig. 2.8).

We also compared the size of the segmented breast area in the previous view with the size in the current view. This resulted in a scale factor which was on average also lower than one. However, the linear relationship between both scale factors turned out to be weak. Large differences in compression between previous and current view, resulted in both a large MI scale factor and a large size scale factor. However, relatively large size scaling factors could sometimes also be explained by difficulties with segmenting the breast contour. On high contrast films the skin air contour is sometimes not visible due to the limited optical density range of our digitizer. Large MI scale factors were selected when there were large positioning differences between the temporal views and sometimes to correct for the reduction of glandular tissue.

A potential problem with MI when large developing densities are present was investigated. It appeared that MI is not hampered by the changes between the previous and the current view. However, further study on a larger data set will be needed to investigate the robustness of MI with respect to large developing densities.

We compared the performance of the registration methods for both MLO and CC image pairs. The distance in the image matrix between annotations in temporal pairs turned out to be smaller for lesions in CC views than in MLO views both before and after registration. Only for MI registration, median distance errors turned out to be equal for lesions in MLO and in CC views. For warping and CM registration, this effect might be caused by the inaccuracy of the pectoral muscle segmentation. However it is more

likely that this effect is due to variations in positioning. It is well known that even for well trained radiographers the positioning varies between subsequent views. As the MLO view positioning is more difficult, it would not be surprising that even for the unregistered views there is more variation for the MLO images.

Our results could be compared with those obtained by the regional registration technique developed by Hadjiiski *et al.* (2001a). They developed a multi-stage regional registration technique and tested it on a set of 124 temporal pairs obtained with an average time interval of one year. In the first stage a fan-shaped search region is defined on the prior mammogram based on the location of the lesion of interest in the current mammogram. As the basis of this fan-shaped region the nipple location was chosen, which was determined manually. Radial and angular size of this search region were experimentally chosen based on the used data set. After this initial stage they found an average distance error of 8.4 mm, which could be reduced to 4.2 mm with subsequent regional matching stages. Using our global MI registration approach we found an average distance error of 7.9 mm. We did not investigate local matching to improve lesion registration. It is noted that the advantage of our technique is that it is fully automatic, and did not require manual nipple segmentation. Comparison with other studies was not possible, because these did not include a validation step.

In conclusion, in this study MI appeared to be the best technique for global mammogram registration. It seems that the average distance error obtained with this technique is sufficient for display purposes and the method does not rely on accurate segmentation of the breast or its internal structures. In CAD applications this global registration may facilitate detection and analysis of temporal lesion pairs. In the current study the distance between annotations in two subsequent temporal views is used as a criterion to test global mammogram registration. Due to deformation of the breast during the imaging procedure it is very difficult to obtain more accuracy with global registration. If needed, a regional registration step can be used in addition, for instance to correct for slight rotation of a lesion.

Chapter 3

Optimized perception of lesion growth in mammograms using digital display [1]

3.1 Introduction

In mammography, it is generally believed that growth of lesions is an important clue to detect lesions and to discriminate between benign and malignant. However, in a literature review it appeared that the importance of obtaining previous mammograms to compare with new studies is not clear. Bassett *et al.* (1994) studied the usefulness and costs of comparing previous and current mammographic examinations. They concluded that comparison with previous mammograms has a positive effect on cancer detection and clinical management, however the overall costs and time involved in obtaining previous mammograms are substantial. Thurfjell *et al.* (2000) evaluated the effect of old mammograms on the specificity and sensitivity of radiologists in mammography screening. They found a significant increase in the specificity, while the effect on the sensitivity was unclear. Callaway *et al.* (1997) also studied the effect of previous films on mammographic interpretation and detection of breast carcinoma. In contrast to Thurfjell, they did not find an improvement in diagnostic accuracy, only a significant reduction in additional examinations that would have been requested. Wilson *et al.* (1996) performed a retrospective evaluation in women with mammograms that were initially interpreted as negative or negative with benign findings pending comparison with a mammogram from a previous examination. In only 1 % of the cases comparison with the previous mam-

[1] The content of this chapter has been published previously in van Engeland *et al.* (2003b).

mograms altered the original interpretation. Recently, Burnside *et al.* (2002) published a study to analyze the differential value of comparison with previous examinations. They found that for screening mammography comparison with previous examinations significantly decreases false-positives, but has no effect on true-positive findings. They also mention that comparison permits detection of cancers at an earlier stage, which seems to be in contradiction with the previous statement.

Indirect evidence for the fact that temporal change analysis is important can be found in papers presenting studies about missed cancers, follow-up of probably benign lesions and comparison of initial and subsequent screening. Bird *et al.* (1992) analyzed the characteristics of missed lesions and compared them with correctly diagnosed cancers. One of the results they found was that missed cancers were more likely to demonstrate a developing opacity as an indication of cancer. Sickles (1991) found 17 malignant cases in a set of 3,184 consecutive cases of non-palpable, probably benign breast lesions, for which the majority manifested by means of interval mammographic change. Frankel *et al.* (1995) studied the difference between initial and subsequent screening mammography. Next to the expected effect of finding less and smaller lesions for subsequent screening, they also found a significantly higher biopsy yield for subsequent screening, demonstrating the positive effect on the radiologists' interpretation when using prior studies.

All studies investigating the usefulness and costs of using previous mammograms question the benefit on diagnostic accuracy because of the high costs of acquiring the old mammograms (Wilson *et al.* 1996; Bassett *et al.* 1994). However, in the near future digital archiving of mammograms will reduce the costs and the efforts needed to acquire old mammograms.

Digital display of mammograms also offers new ways to optimize display that may lead to improved perception of lesion growth. Introduction of digital mammography also opens new ways to design display strategies for screening mammography. Optimized display might lead to improved perception of mammographic changes over time. In this study we investigated two ways of presenting prior and current mammograms: next to each other (*simultaneous*) and alternating at the same display (*toggle*). The mammograms were displayed on a dedicated mammography workstation. The stimuli were created using extracted lesions that were pasted into normal mammograms. For each stimulus the observer was asked to select the image containing the largest lesion. Threshold detection levels of lesion growth were determined under both conditions.

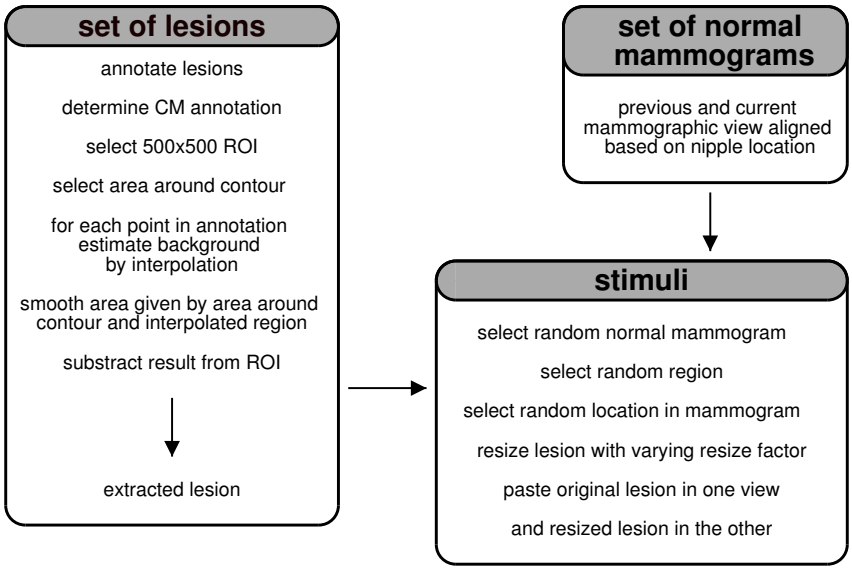


Figure 3.1: *Creating the stimuli.*

3.2 Methods

3.2.1 Creating the Stimuli

The process to create the 210 stimuli consisted of three main steps (Fig. 3.1). First a set of extracted regions was created. Next, a set of normal mammograms was selected. And finally, the stimuli were created by pasting the lesions into the normal mammographic views. These steps are described in the following paragraphs.

Creating the set of extracted lesions

Lesions were annotated in 100 μm images (Fig. 3.2). Regions of interest (ROI's) of 500 $\times$ 500 pixels containing these lesions were selected, centered at the center of mass of the annotation. The background of each lesion was estimated by interpolating the intensities on the annotation contours. The regions around the annotations and the interpolated areas were then smoothed and subtracted from the original ROI's. In the experiment the lesion set consisted of 7 extracted lesions.

Acquiring the set of normal mammograms

Two normal mammograms of mostly fatty breasts were selected for which images of two subsequent screening rounds were available. The previous and current mediolateral oblique (MLO) views of the right breast were used and aligned based on the nipple position.

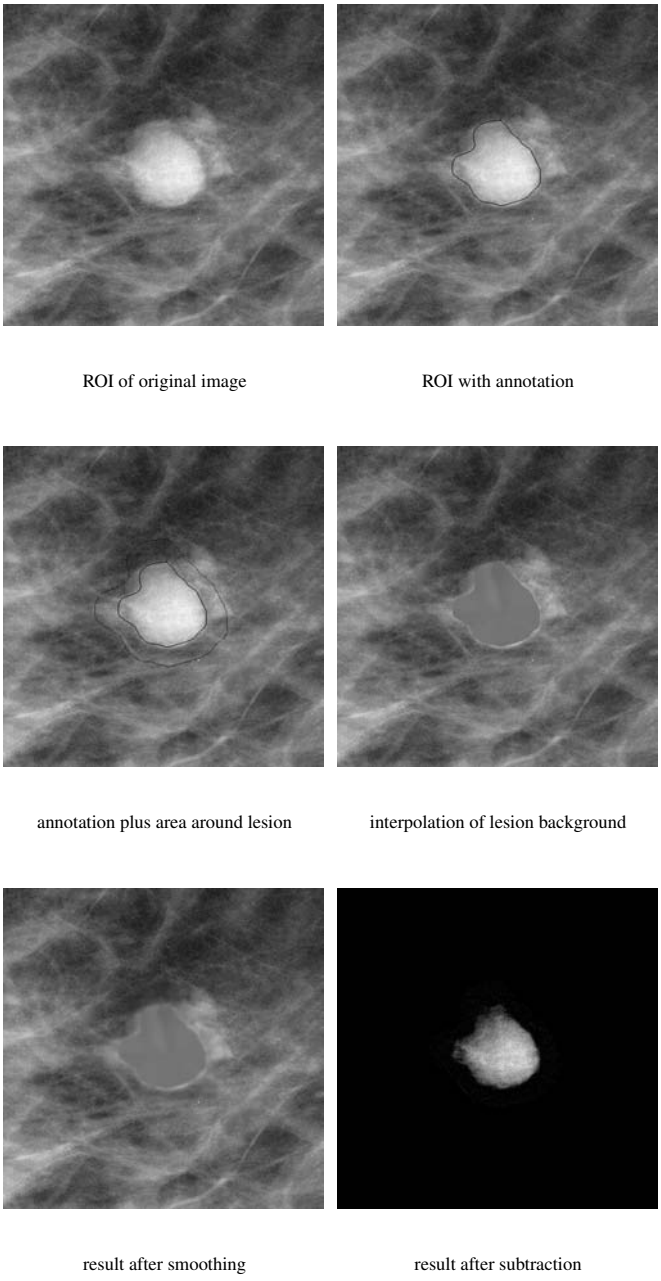


Figure 3.2: *Extracting the regions. See also text.*

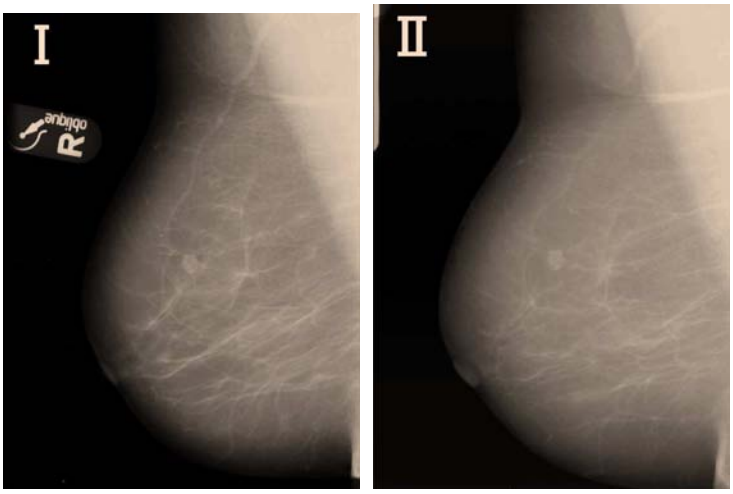


Figure 3.3: *Example of a stimulus. The lesion diameter on the current view (II) is 5 % larger than the lesion diameter on the previous view (I).*

Final stimulus creation

For every stimulus one of the normal mammograms was selected. Next, a lesion was selected at random from the set of extracted lesions. A random location in the previous mammogram was selected and it was checked whether this was a suitable position to paste the lesion, i.e. inside breast area, not too close to skin or pectoral muscle and not in a high intensity region. When the location satisfied these conditions, the lesion was resized using a selected resize factor and bilinear interpolation. The original lesion was then pasted into the previous view, and the resized lesion was pasted into the current view at the same location, or the other way around, making sure that the final set contained as many stimuli where the largest lesion was in the previous view as in the current view. When the lesion location did not fulfill the requirements another location was selected until success. Seven resize factors were used ranging from 1.0 to 1.1.

3.2.2 Experimental setup

The stimuli were displayed on a dedicated mammography workstation (MBC-SCR1, MeVis BreastCare) that was developed by a European Consortium in which we partici-

pated [1]. The workstation was equipped with two high-resolution CRT displays (BARCO, MGD 521, 300 Cd/m^2, using BarcoMed 5MP1H 12 bit graphics boards). The used workstation was a research prototype allowing swapping of images on the display with a transition time of 0.52 ± 0.01 seconds.

In a 2AFC paradigm, the observers were asked to select the image containing the largest lesion. Figure 3.3 shows an example of a stimulus. In half of the trials, the previous and current view were displayed next to each other (*simultaneous*). In the other half of the trials the observers could alternate between the previous and current view (*toggle*). The toggle- and simultaneous-trials were presented in mixed order. For both conditions 210 trials were presented, resulting in 30 observations per resize factor per condition.

Four observers participated in the experiment. The observers were physicists working in the field of mammography. The conditions were similar to standard reading conditions for screening mammography. There was no time limit and no limit on the number of alternations for decision making. The observers used two sessions of approximately one hour each. Time intervals of images displayed in toggle mode used by the observers typically ranged from 0.5 to 1.5 seconds.

3.2.3 Evaluation

The results of the experiments are presented in the form of psychometric curves, representing the relationship between stimulus level and observer response. In our experiments the stimulus level was a measure of the size difference between the lesions on both views. The psychometric curves show the fraction of correct observer responses as a function of the resize factor. A logistic psychometric function (Eq. 3.1) was used.

$$\psi = \frac{1}{1 + e^{(\alpha - x)/\beta}} \tag{3.1}$$

where, ψ is the fraction of correct responses, x is the resize factor, α is the threshold, and β determines the slope.

The experimental setup was such that α is one. The free parameter β was fitted using MUEST (Snoeren & Puts 1997).

3.3 Results

In Fig. 3.4 the results for all four observers are presented. The graphs show that observers 2-4 performed more accurate in selecting the largest lesion when using the toggle option. For observer 1 there was no difference in performance.

[1] Soft-Copy REading ENvironment SCREEN, EU-Project, IST-1999-10246.

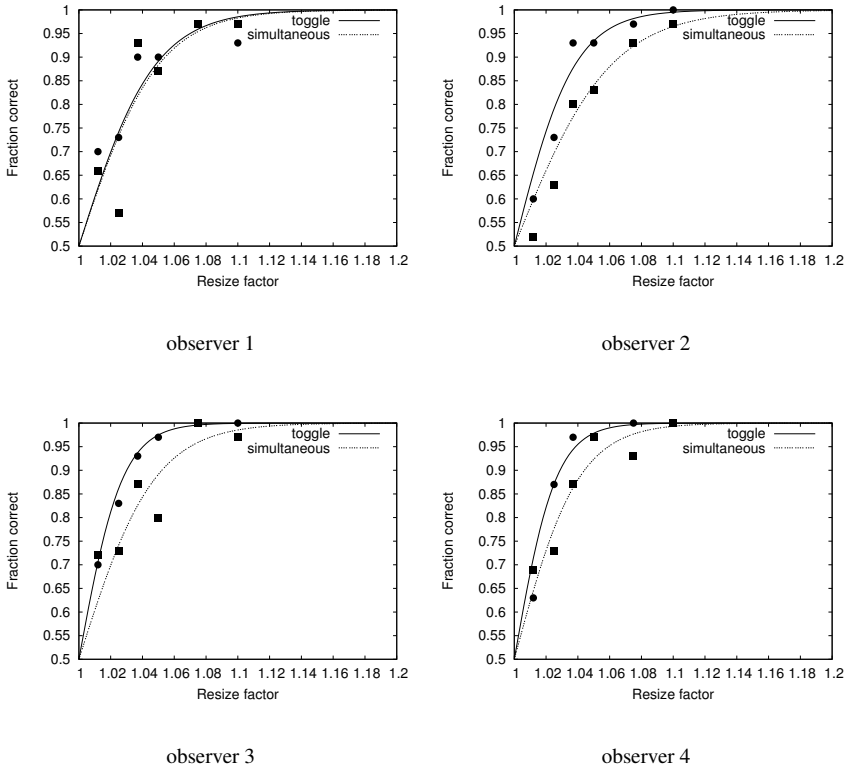


Figure 3.4: *Psychometric curves for all four observers.*

Table 3.1 shows a statistically significant increase in slope of the psychometric curves when comparing simultaneous display and toggling (paired t-test, one-tailed, P = 0.027). Table 3.2 presents the thresholds for 90 % correct observer response.

So far, results have shown an increase in performance when the observers used the toggle option. Since a significant increase in decision time when using the toggle option would be a drawback for this technique during screening, we also investigated the decision times.

For three of the four observers the time needed to decide which image contained the largest lesion, was less when displaying the images simultaneous (Fig. 3.5). This difference between decision time for toggle and simultaneous conditions was however not statistically significant (P = 0.09, Table 3.3).

Table 3.1: *β values (slope at x = 1.0 is 0.25$*\beta$) of psychometric curves for simultaneous and toggle condition for all four observers.*

	simultaneous condition	toggle condition
observer 1	40.5 ± 6	42.1 ± 6
observer 2	33.2 ± 5	53.1 ± 7
observer 3	42.1 ± 6	73.1 ± 11
observer 4	49.8 ± 7	74.4 ± 11

Table 3.2: *Size change detection thresholds for 90 % correct observer response.*

	simultaneous condition	toggle condition
observer 1	5.4%	5.2%
observer 2	6.6%	4.1%
observer 3	5.2%	3.0%
observer 4	4.4%	2.9%

Table 3.3: *Median decision time (sec) for both the simultaneous and the toggle condition per observer.*

	simultaneous condition	toggle condition
observer 1	8.25	9.84
observer 2	12.48	11.55
observer 3	10.06	12.80
observer 4	9.94	13.80

3.4 Discussion

Interval change analysis is considered to be very important for both lesion detection and classification into benign and malignant. We have shown that observers perform more accurately in detecting lesion growth when alternating at the same display (toggling). Although, the difference in performance between both display methods was significant, the improvement in size change detection thresholds might seem rather small. However, even a small improvement in size change detection can be important, since mammography screening is particularly useful in detection of slow growing tumors. Furthermore, observers mentioned that they were hampered by the fact that during toggling there was a

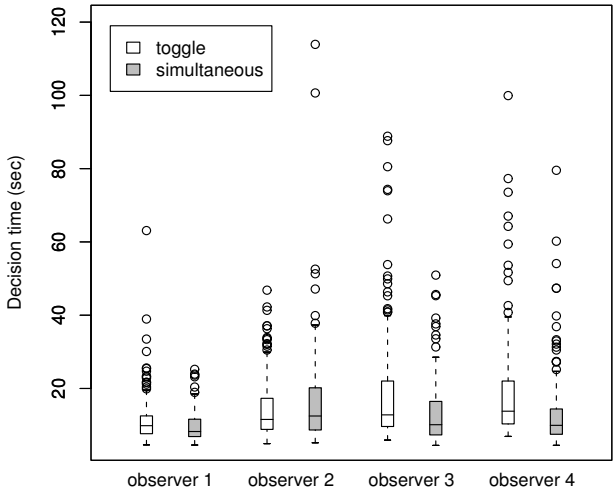


Figure 3.5: *Decision times for each observer for toggle and simultaneous conditions. The median values, upper and lower quartile values are plotted. The whiskers extend to the most extreme data point which is no more than 1.5 times the length of the box away from the box, and the circles represent the outliers.*

short period of time when there was no image on the screen. We expect that shorter image transition times will improve observer performance, and as a consequence increase the difference between toggling and displaying simultaneous. A shortcoming of our study might be that no radiologists participated in the experiment. However, since it is not necessary to have extensive knowledge of mammography for this particular task of growth detection, we do not expect that the outcome of the experiments would be different for a group of radiologists. Besides, all our observers worked in the field of mammography, and were therefore very familiar with mammograms.

In conclusion, in the near future digital mammography will reduce costs and efforts to obtain previous mammograms, and digital display can lead to improved perception of lesion growth. Therefore it is expected that the benefit of comparison with previous mammograms will outweigh the costs and efforts needed for the retrieval of previous studies.

Chapter 4

Regrouping initial CAD mass detections to facilitate classification of suspicious regions in mammography [1]

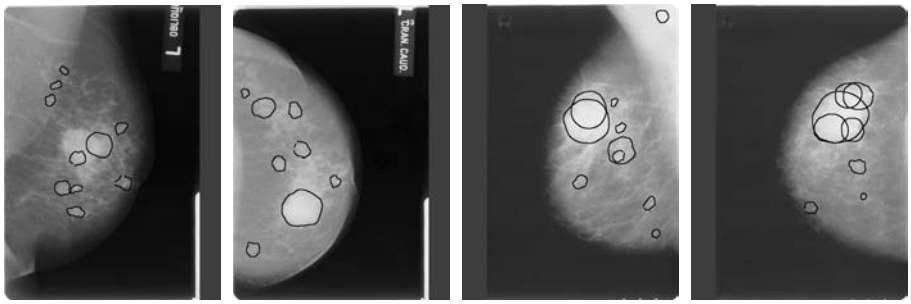
4.1 Introduction

Most development of CAD for mammography has been based on the analysis of single views. By combining information from multiple mammographic views (temporal, mediolateral oblique (MLO) and craniocaudal (CC), or bilateral) it should be possible to improve the accuracy of CAD methods.

In literature some approaches have been described to establish correspondence between multiple views. However, few studies have been conducted to use the multiple view information to improve classification. Hadjiiski *et al.* (2001b) developed interval change analysis of masses and found that their classification accuracy improved significantly. Paquerault *et al.* (2002) developed a two view matching method which results in a correspondence score for each possible mass pair. By combining this correspondence score with their single view detection score, their classification results also improved significantly.

During the development of multiple view techniques,we have noticed that these techniques are hampered by the phenomena that lesions are sometimes detected by multiple regions (van Engeland *et al.* 2002). On a data set containing 385 images with a mass le-

[1] The content of this chapter has been published previously in van Engeland & Karssemeijer (2005).



(a) (b) (c) (d)

Figure 4.1: *Examples of initial detections, showing multiple detections on the lesion. The views in (a) and (b) belong to the same case, as do (c) and (d).*

sion we found that our CAD scheme initially detected 20 percent of the mass lesions by multiple regions. This complicates the process of automatically linking lesions visible in multiple views. In addition, it might also decrease single view detection performance, as region based features are not representative for the characteristics of the entire lesion. In Fig. 4.1 CAD results for the MLO and CC image of two cases are presented. For the case presented in Fig. 4.1(a) and 4.1(b) the lesion is detected in the CC view as one region, and in the MLO view it is detected by two non-overlapping regions. For the case presented in Fig. 4.1(c) and 4.1(d) the lesion is detected by several overlapping regions in both views. In such a situation it is not possible to find a one-to-one correspondence between regions in the MLO and CC view. Furthermore, even if correct correspondence is found, combination of feature values into multiple view representation will not be optimal, because some regions will describe only part of the lesion while others will represent the whole lesion. Therefore, we aim at developing a technique in which lesions hit by multiple regions are recognized. Subsequently, they can be re-segmented and re-evaluated as a single region.

The initial step in our CAD algorithm creates a likelihood image that contains for every pixel a malignancy score based on characteristic tumor features, like spiculation. In this likelihood image peak detection is done at one scale. Our detection scheme was optimized for the detection of lesions with a diameter around 15 mm. Multiple peaks on the same lesion can occur for relative large densities with inhomogeneous pixel value distribution. Also when groups of spicules present at some locations on the edge of a lesion do not all point to the same center, this sometimes leads to multiple detections on the same lesion. It is difficult to overcome these problems with a multi-scale approach, since characteristic features like spiculation are only visible for a certain scale. The

possible benefit of using multi-scale detection has been studied before by te Brake & Karssemeijer (1999). Their study showed only minor improvements for the multi-scale method over the single-scale approach, provided that an optimal scale was chosen.

In our current CAD method, an algorithm searches for peaks that are located closer than 8 mm together, in order to avoid multiple detections on the same lesion. If these multiple detections are found the detection with the lowest likelihood value is removed. This is a basic approach that removes many of the multiple detections, however we have found that when correspondences between detections are to be determined in multiple view analysis, this approach did not suffice. The current study describes the development of an algorithm to remove multiple detections that does not only take the distance between initial detections into account, but also at the image structure between detections. When correspondence is found, the two detections are replaced by a new detection in between the initial detections.

After a short description of our CAD scheme (Sec. 4.2.2), the regrouping technique is described in detail in Sec. 4.2.3. Results are presented in Sect. 4.3.

4.2 Methods

4.2.1 Data set

The used data set consisted of 201 cases, containing 804 images. All cases had four-view mammograms. There was a mass lesion visible in 385 images. In 20 percent (79 images) the mass lesion was detected by multiple regions. The mammograms were taken from an annotated database containing cases from the Dutch breast cancer screening program. The mammograms were digitized with a Lumisys 85 digitizer at a pixel resolution of 50 μm, and averaged down to a resolution of 200 μm, maintaining the original gray value depth of 12 bits.

4.2.2 CAD scheme

Our CAD scheme consists of the following steps (Fig. 4.2):

- Segmentation of the mammogram into breast, pectoral muscle, and background area,
- Initial detection step resulting in a number of suspect image locations,
- Region segmentation with dynamic programming using these locations as seed points,
- Final classification step to classify regions as true abnormalities and false positives.

These steps wills be described in more detail in the following paragraphs.

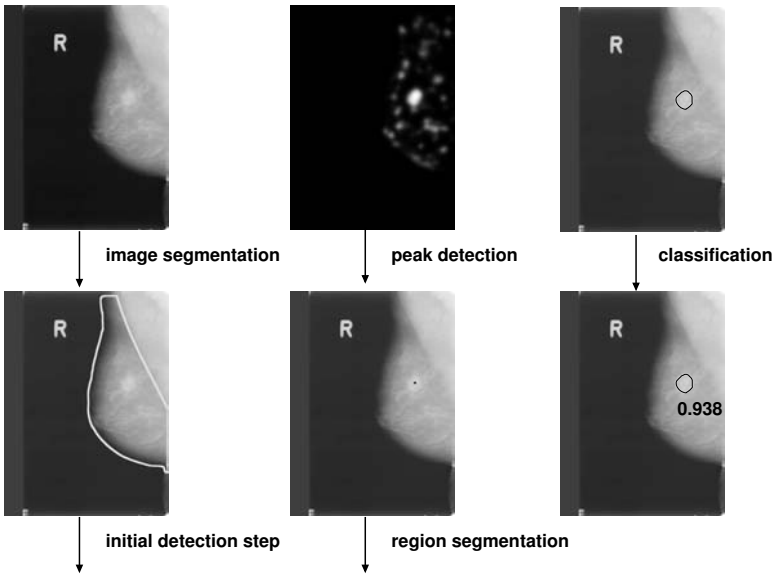


Figure 4.2: *Overview of our CAD scheme.*

Segmentation of the mammogram Our breast segmentation algorithm, which is similar to the method described by Highnam & Brady (1999), is based on the idea that the background is very flat and can be identified by low gradient values. We use a fixed threshold on the gradient magnitude to determine pixels in the background. In addition, a threshold on the image intensity values is used to exclude brighter pixels from the background. This threshold is determined by histogram analysis. Morphological operations are applied to merge regions in the breast and to remove labels.

The pectoral muscle segmentation technique is described in detail in Karssemeijer (1998). It selects a region of interest in the image where the pectoral muscle is likely to be located. By applying the Hough transform on this region a parameter space, which represents straight lines through the region, is obtained. Finally, the highest peak is selected, which represents the straight line estimate of the pectoral muscle boundary. To improve the segmentation, pixels with a stronger gradient have a larger increment in the Hough accumulator, and a correction is made to reduce the bias towards longer pectoral muscle boundaries. Furthermore, if there are more candidate peaks in Hough space, the peak with the largest corresponding pectoral muscle area is selected.

Initial detection step For each image pixel inside the breast area a number of features are calculated that represent tumor characteristics. These features are related to the presence of spicules (Karssemeijer & te Brake 1996), and to detecting a central mass (te Brake & Karssemeijer 1999). A neural network is used to classify each pixel using these features and assigns a measure of suspiciousness to it, resulting in a so-called likelihood image $L(i, j)$. This likelihood image is smoothed, and a peak detection is performed at one scale. The peak detection algorithm visits every pixel in the likelihood image and a peak is detected when the likelihood is above a certain threshold and there are no other pixels in the neighborhood (square neighborhood with size 32 x 32 pixels) with a higher likelihood value. This results in a number of suspect image locations. In our current CAD method, an algorithm searches for peaks that are located closer than 40 pixels together, in order to remove multiple detections on the same lesion. If these multiple detections are found the detection with the lowest likelihood value is removed.

Region segmentation The initial detection step results in a number of suspect image locations. Each of the detected peaks is used as seed point for region segmentation, based on dynamic programming (Timp & Karssemeijer 2004).

Final classification For each region, features are calculated that describe the position of a region in the breast (for instance the distance to the pectoral and the skin), region size, contrast, texture, compactness and acutance measures. A neural network is used to classify regions as true abnormalities and false positives based on these features.

4.2.3 Regrouping algorithm

The algorithm starts with a list of locations that are sorted based on the likelihood value of each location, where the location with highest likelihood value is first in the list. The algorithm then checks for the first location whether there is another detection within a distance of 120 pixels (24 mm). If this is the case, dynamic programming is used to find the path with the highest pixel level between these detections. The path finding algorithm is described in detail in the next section. Next, the variance of the pixel values along this path is calculated and normalized with the pixel value variance in the whole breast area (see Sec. 4.2.5). If this value is below a certain threshold the two detections are joined by removing the location with the lowest likelihood value. The detection with the highest likelihood is replaced by a new detection in between the combined detections. Then the algorithm searches further to find other detections that can be combined with this detection. The same procedure is followed for the remaining detections in the list. We have set a limit to the number of detections that can be combined. After the algorithm is finished a combined detection can consist of two to four initial detections. Fig. 4.3 describes the algorithm in pseudo-code.

```
for   (k = 0; k < ndetections - 1; k++)   {
   joined = 0
   for   (l = k+1; l < ndetections; l++)   {
         if   (distance(k,l) < 120 && detection[k].valid > -1 &&
            detection[l].valid > -1 && joined ≤ 1 && varpath(k,l) < 0.04 &&
            detection[k].level > 125 && detection[l].level > 125)   {
               joined++
               detection[l].valid = -1
               detection[k].x = (detection[l].x - detection[k].x) / 2 + detection[k].x
               detection[k].y = (detection[l].y - detection[k].y) / 2 + detection[k].y
               detection[k].level = (detection[k].level + detection[l].level) / 2
         }
   }
}
```

Figure 4.3: *Regrouping algorithm in pseudo-code. Where $ndetections$ is the number of initial detections, $distance(k,l)$ is the distance between the k-th and the l-th detection, $valid$ indicates whether a detection is still active and has not yet been discarded, $varpath(k,l)$ is the normalized variance of pixel values along the path from the k-th to the l-th detection, $level$ indicates the likelihood value for a detection, and $joined$ indicates whether a detection has already been used in a combination before. This last variable is used to ensure that not too many initial detections are combined into one.*

4.2.4 Finding the highest pixel value path between two detections

Once the algorithm has found two detections, k and l, close together, it searches for the path between k and l with the highest pixel value using dynamic programming. This procedure starts with the selection of a region of interest (ROI) containing both detections. A rectangular region is formed by resampling the image perpendicular to a line connecting k and l. The length of this ROI is equal to the distance between the detections and the height is set to 50 pixels. Next, a cost matrix is generated, where the costs per pixel are defined as:

$$c(i,j) = \frac{max(g) - g(i,j)}{max(g)}, \tag{4.1}$$

where $max(g)$ is the 99th percentile of the pixel values in the ROI. The first column in the cost matrix represents the start nodes for the algorithm, whereas the end nodes are represented by the pixels in the last column of the cost matrix. The cumulative cost matrix of each path stores the cumulative costs for each path. In the first column the

pixels are set equal to the cost of these pixels:

$$C(i,0) = c(i,0), \tag{4.2}$$

where $C(i,0)$ is the cumulative cost, and $c(i,0)$ is the cost value for pixel $(i,0)$. The other pixels in the cumulative cost matrix are calculated by a recursive step:

$$C(i,j+1) = \min_{-m \leq l \leq m} C(i+l,j) + c(i,j+1) + h(l), \tag{4.3}$$

where l is the direction, and $h(l)$ is the cost assigned to direction l. For searching the maximum pixel value path we have set the direction costs to zero. The end point of the path is the pixel in the last column of the cumulative cost matrix with the lowest cost. The optimal path is found by back tracing from the end pixel to one of the pixels in the first column. Since the detections are peaks in the likelihood image and not in the pixel value image, the begin and end points of the path do not have to coincide with the detections k and l.

4.2.5 Estimating pixel value variance along path

The variance in pixel values along the path between two detections is used to decide whether two detections should be combined. The variance along the path is normalized with the variance of the pixel values in the breast area of the image, excluding the pectoral muscle. The variance of the pixel values within the breast area of the image is given by:

$$V_{image} = \frac{\sum_{(i,j)\in B}(g(i,j) - \bar{g})^2}{N-1}, \tag{4.4}$$

where N is the number of pixels inside the breast area B, and $\bar{g}$ is the average pixel value. And the relative pixel value variance V_{path} along the path is given by:

$$V_{path} = \frac{\frac{\sum_{(i,j)\in P}(g(i,j) - \bar{g_p})^2}{Np-1}}{V_{image}}, \tag{4.5}$$

where P is the set of pixels on the path between k and l, $\bar{g_p}$ is the average pixel value on the path, and Np is the path length.

Figure 4.4 shows the relative variance values along the paths between false positive (FP) detections, true positive (TP) detections, and TP and FP detections.

The threshold value for the path variance was experimentally chosen, and for $var(path) < 0.04$ detections were considered as belonging to the same structure and were combined.

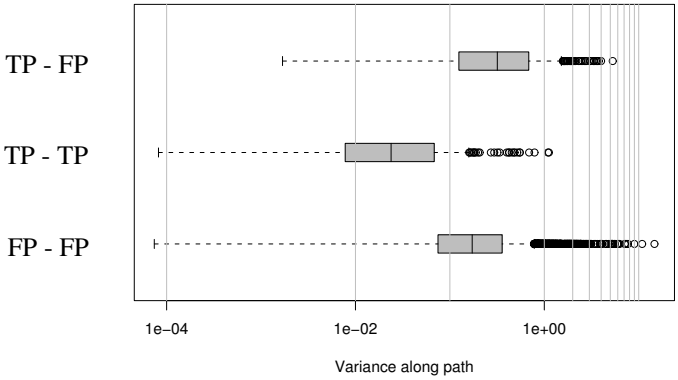


Figure 4.4: *Relative variances of pixel values along paths between detections for different types of region combinations. Boxes show lower quartile, median and upper quartile values. The whiskers indicate the minimum and maximum variances.* ○ *represent outliers, which are data points with a distance to the upper quartile value of more than 1.5 times the interquartile range.*

4.2.6 Evaluation

The regrouping algorithm was evaluated in two ways, by determining the number of correctly linked TP - TP initial detections and the number of erroneously linked TP - FP detections.

Also, the algorithm was evaluated both by comparing the number of incorrect links established between MLO and CC regions and by comparing the detection performances, with and without applying the regrouping algorithm. The linking algorithm to find links between regions in corresponding MLO and CC views is described in short in the following section, and Sec. 4.2.6 describes the evaluation of the classification between TPs and FPs.

Region correspondences in MLO and CC view

The initial goal of this study was to try to reduce the number of incorrect links between MLO and CC regions for our multiple view detection scheme. For each region in a mammographic view our linking algorithm searches for the most likely region match in the corresponding view. The most important feature that is used is the distance from a region to the nipple, since it is generally believed that this distance remains fairly constant

between views. Other features that are used to find corresponding regions in MLO and CC view are: region size, contrast, and spiculation. To determine the likelihood of links we determine a cost function which assigns a higher cost to a link when features do not match. The best possible link is found by searching for the link with the lowest cost. The cost function is defined as follows:

$$C(p,q) = \sum_{f \in F} 2 * (-0.5 + \frac{1}{1 + e^{-\frac{|f_p - f_q|}{0.5 * (f_p + f_q)}}}), \tag{4.6}$$

where F is the set of features used, p and q are regions on corresponding views, and f_p is the feature value of feature f on region p.

For this study parameter settings and the choice of combination features was not fully optimized. Our main concern was to test whether the regrouping technique resulted in less cases with wrong MLO - CC region combinations. Region combinations could only be tested for true positive regions, using annotations made by a radiologist.

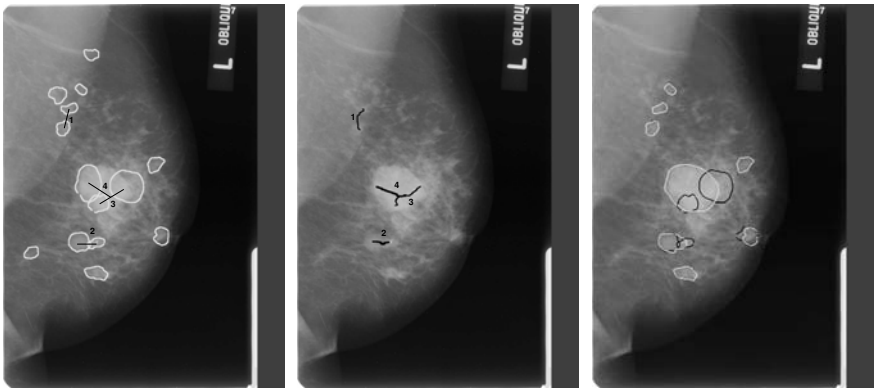
Detection performance

The detection performance after regrouping of suspicious locations is compared with the original detection performance on all mammograms in the data set. A three-layer backpropagation neural network is used with 8 hidden nodes. The classifier is tested using crossvalidation with 90 percent training and ten percent testing. Free-response Receiver Operating Characteristic (FROC) analysis is used to present the results.

4.3 Results

4.3.1 Example image

In this paragraph the use of the regrouping algorithm is demonstrated on the image in Fig. 4.1. Figure 4.5(a) shows the initial detections, represented by their region segmentations. Also four possible region combinations are shown. Combinations 1 and 2 are FP - FP region combinations and 3 and 4 are TP - TP combinations. For all four possible region combinations the paths are determined with the highest pixel values (see Sec. 4.2.4). These paths are presented in Fig. 4.5(b). Figure 4.6 shows the pixel values and likelihood levels along each path. After calculation of the variances of the pixel values along the paths, the algorithm decides to combine the detections attached by path 1, 3 and 4. This can also be appreciated from the plots in Fig. 4.6, although paths 3 and 4 show a minimum in the likelihood value, leading to the detection of two separate locations, the pixel values are relatively constant. The result after combination of detections and segmentation using the new seed points are presented in Fig. 4.5(c).



(a) (b) (c)

Figure 4.5: *Example image showing initial detections. Four detection combinations are selected and used to demonstrate the path finding procedure (a). Paths with highest pixel values between initial detections found by dynamic programming (b). Result; original detections (black contours) and regrouped detections (white contours) (c).*

4.3.2 Results of regrouping

Table 4.1 summarizes the combination results for the whole data set. In 48 percent (38 of the 79 multiple region lesions) the initial detections were correctly combined. For two percent of the FPs in the data set a combination was established with another FP (106 out of 5891 FP detections). In one percent (5 of the 385 lesions) a TP region was incorrectly combined with a FP region.

Table 4.1: *Combination results, number of different types of combinations established.*

	FP - FP	TP - TP	TP - FP
no. of links	106	38	5

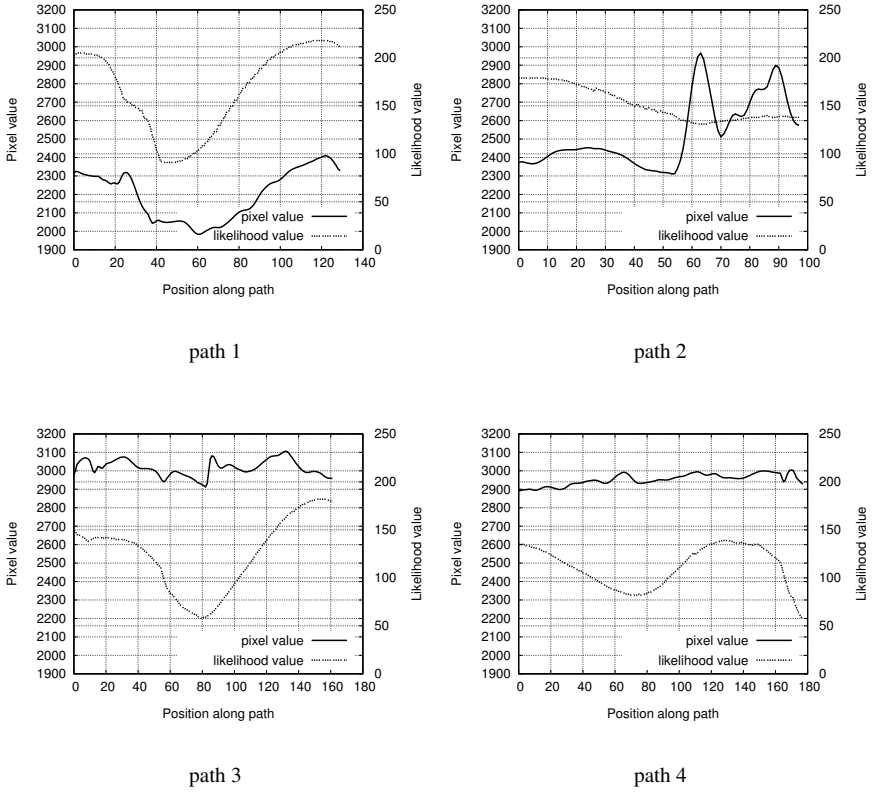
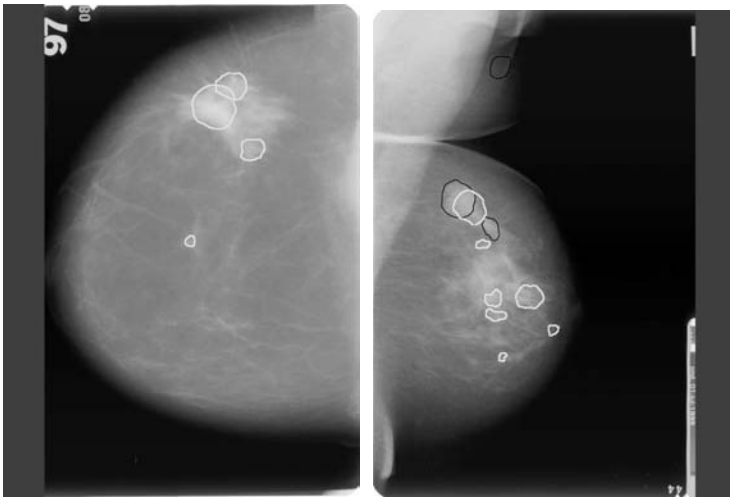


Figure 4.6: *Pixel values and likelihood values along paths between detections in Fig. 4.5. Note that for the second path the variance was calculated only with pixels not containing calcifications.*

At its current stage of development, the regrouping algorithm is not able to establish a link between initial detections on the lesion in half of the multiple region cases. The reasons for these missed TP - TP combinations are (in order of importance):

- the lesion looks very diffuse on the mammogram (see for example Fig. 4.7(a))
- the exact location and size of the lesion is unclear on the mammogram, for instance in case of a invasive lobular carcinoma, which is a type that may be harder to detect in a mammogram than the common invasive ductal carcinoma,
- a large part of breast is affected, resulting in many CAD regions over a large area of the breast,



(a) (b)

Figure 4.7: *Example where the initial detections are not combined in an image with a diffuse lesion (a). Example of a wrong initial detection combination (b). Black contours indicate the original detections, white contours indicate the regions after regrouping.*

- the boundaries of the lesion are not clear,
- the lesion consists of more parts, also microcalcifications were annotated, and
- the lesion is visible as an architectural distortion.

Reasons for incorrect TP - FP combinations can be summarized as:

- the exact location and size of the lesion is unclear on the mammogram,
- the lesion is located in dense glandular tissue and boundaries are not clear, and
- there is an elongated structure attached to lesion (see for example Fig. 4.7(b)).

4.3.3 Region correspondences in MLO and CC view

The application of the regrouping algorithm reduced the number of cases with incorrect links from 70 to 62 (out of 201). Of the remaining incorrectly linked cases a large part, 30 percent, could still be attributed to multiple detections on the lesion. However, the major

reason for incorrect region links were caused by large differences in positioning between the CC and MLO view and unclear nipple position on one or both views. Both these situations make the distance between the nipple and the lesion a less reliable measure for linking.

4.3.4 Detection performance

Figure 4.8 shows the detection performances of our CAD scheme with and without using the regrouping algorithm. Both image and case based performances are provided. Both image and case based curves show an improvement in the range of 0.02 - 0.3 FP/image. However, statistical significance of this improvement could not yet be demonstrated.

4.4 Discussion

The results presented in Sec. 4.3.2 show that the regrouping algorithm correctly combined the initial detections in half of the cases with multiple initial detections on the lesion.

When running our linking algorithm to link regions in corresponding CC and MLO views, we found that application of the regrouping algorithm reduced the number of cases with incorrect links by 10 percent.

Incorporation of the regrouping algorithm in our CAD scheme resulted in a slight improvement of the detection results. This effect might be caused by the the fact that application of the regrouping algorithm reduced the number of lesions detected by multiple regions, leading to better region based features. In addition, recombination of FPs reduces the number of FPs in the set, which might also have attributed to the improvement in the detection performance. However, since the number of combinations is relatively small compared to the total number of initial detections in the data set, this improvement in detection performance was relatively small.

At its current stage of development, the regrouping algorithm is not able to establish a link between initial detections on the lesion in half of the multiple region cases. The major reason for the missed cases is that some lesions appear very diffuse on the mammogram. This makes it more difficult for our algorithm to combine the initial detections. Also for some lobular carcinomas, which are in general harder to find in a mammogram than the more common ductal carcinoma, the algorithm failed to establish a link between the initial detections. Next to the correct initial detection combinations, the algorithm unfortunately also made 5 TP - FP combinations. The majority of these incorrect combinations were established as a result of unclear lesion boundaries, for instance when a lesion is located inside the dense glandular tissue. In two cases, a region on an elongated strand of tissue attached to the lesion, which was not annotated as being part of the lesion,

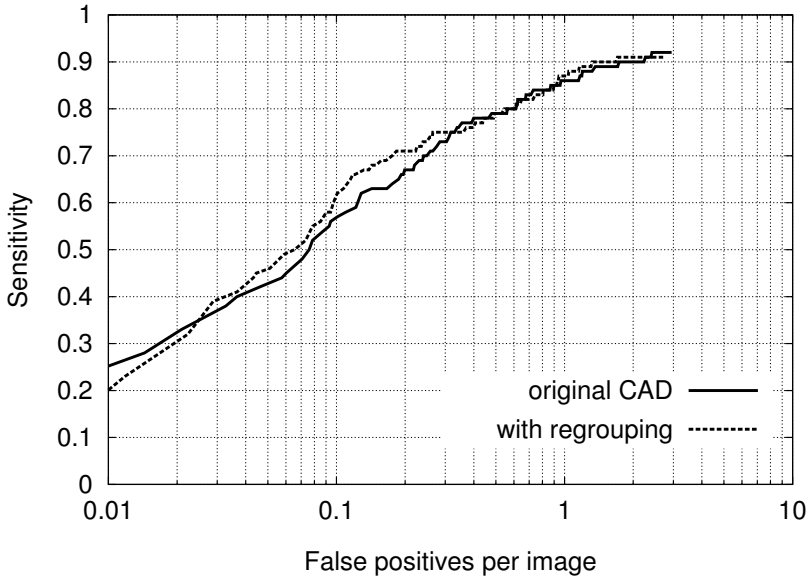
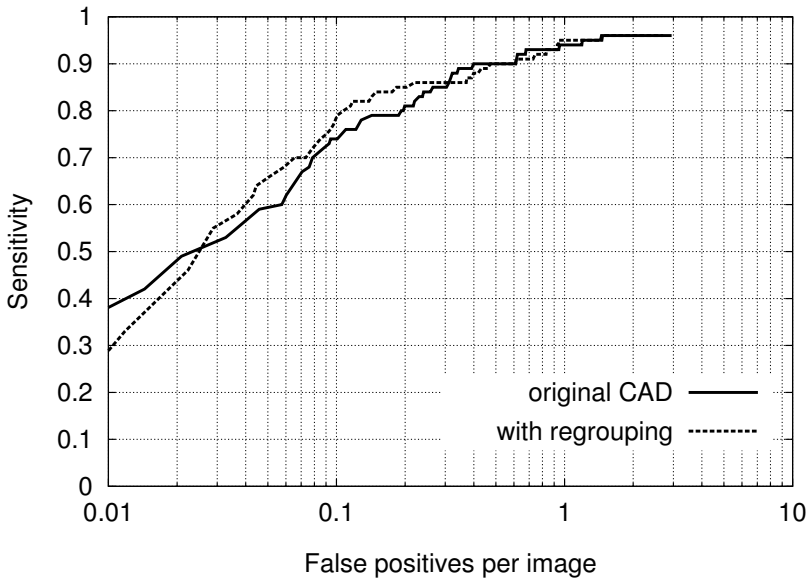


image based



case based

Figure 4.8: *FROC curves comparing the original detection results and the detection results after application of the regrouping algorithm.*

was combined with the region on the actual lesion.

In order to improve the current regrouping algorithm a solution has to be found for the diffuse lesions. It is not only very hard to establish correspondence in multiple views for these diffuse lesions. Also once correspondence is established it is very difficult to find a good feature description of the entire region. Another point of research might be to look closer at the segmentation of the new regions found after regrouping. Segmentation results might improve when changing settings for preferred region size and shape for these larger regions in our dynamic programming algorithm.

4.5 Conclusions

We have developed a technique to regroup initial CAD regions to facilitate the final classification of suspicious regions. The regrouping technique searches for detections that belong to the same structure. To this end, it takes into account the distance between the detections and the image structure along the path between the detections. When correspondence is found, the two detections are replaced by a new detection in between the initial detections. Our regrouping technique correctly regrouped the detections in 48 percent of the masses initially detected by multiple regions. Of the FP detections two percent were combined, and the percentage of TP - FP combinations was one. The results show an increase in detection performance, and a decrease in the number of incorrectly linked regions in corresponding mammographic views.

Chapter 5

Finding corresponding regions of interest in mediolateral oblique and craniocaudal mammographic views [1]

5.1 Introduction

To decrease the number of missed cancer cases during breast cancer screening, Computer-aided Detection (CAD) systems have been developed. These systems are intended to aid the radiologist by prompting suspicious regions. The benefit of CAD for the detection of microcalcifications has been demonstrated by for instance Freer & Ulissey (2001). They showed in a prospective study that the sensitivity of screening increased by 19.5 % by using CAD. This effect was mainly due to the increased detection of microcalcifications. Although some more recent studies also report improved detection due to computer aided detection of masses (e.g. Cupples *et al.* (2005)), the effectiveness of computer aided detection of masses is not undisputed. Many radiologists feel that the performance of CAD for detection of masses should be improved to make it more useful.

Most development of CAD systems has been based on the analysis of single views. To decrease the number of false positives and to improve consistency, there is a lot of interest to develop CAD techniques that use multiple view information (temporal, bilateral or two views of the same breast). Radiologists in breast cancer screening are trained to use comparisons of the left and right breast to identify suspicious asymmetric densities.

[1] The content of this chapter has been accepted for publication in Medical Physics.

Views from previous screening rounds are used to detect developing densities. It is also known that screening with two mammographic views, mediolateral oblique (MLO) and craniocaudal (CC), improves the detection accuracy of abnormalities in the breast, which can be explained by the fact that two projections allow better estimation of conspicuity of lesions and may reveal lesions hidden by glandular tissue in one of the projections (Bassett *et al.* 1987; Thurfjell 1994; Blanks *et al.* 1999).

Such multiple view CAD techniques require a registration step to find corresponding candidate regions in all available views. In literature some approaches have been described to establish correspondence between multiple views. Timp *et al.* (2005) and Sanjay-Gopal *et al.* (1999) developed an automated regional registration technique to identify corresponding lesions in temporal pairs of mammograms. Karssemeijer & te Brake (1998) and Lau & Bischof (1991) both created a mapping between the left and right breast by using a set of control points defined on the skin line, in a method for asymmetry detection. Highnam *et al.* (1998) used a model-based method to find a curve in the MLO view which corresponds to the potential positions of a point in the CC view. Good *et al.* (1999) reported a preliminary attempt of matching computer-detected objects in two views by exhaustive pairing of the detected objects and feature classification. Paquerault *et al.* (2002) developed a two view matching method, for the combination of MLO and CC information, which resulted in a correspondence score for each possible mass pair. By combining this correspondence score with their single view detection score, their detection results improved significantly.

In this paper we present a method to link potentially suspicious areas determined by a CAD scheme in MLO and CC views. It is more difficult to establish correspondence between MLO and CC views than for temporal image pairs, because the breast is compressed in different ways. There is also lack of invariant landmarks. In practice, only the nipple can be used as a reliable landmark. Radiologists use the distance to the nipple to correlate a lesion in MLO and CC views. It is generally believed that this distance remains fairly constant. We have tested this hypothesis in a previous study (van Engeland & Karssemeijer 2001), and found that in 79 percent of the cases the distance to the nipple does not deviate more than 1.5 cm between both views. In the present study we use an automatically determined nipple location to define an annular search area in the other view. For all possible combinations of candidate regions, features that describe the difference in the radial distance from the candidate regions to the nipple, gray scale correlation between both regions and the mass likelihood of the regions are determined. Next, a Linear Discriminant Analysis (LDA) classifier is used to give a correspondence measure for every possible combination. For every region in the original view the region in the other view with the highest correspondence score is selected as the corresponding candidate region.

We intend to use the linking method to combine information extracted from MLO and CC views in order to improve the mass detection results of our CAD scheme for

masses. The idea is that the confidence that a CAD finding corresponds to cancer may be increased when a finding is visible in both views. The suspiciousness of false positives, on the other hand, may be decreased, because these are usually not correlated between views. In this way we may also increase the number of cancers marked in two views, which has recently been found to be very important in practice, as radiologists appear to pay more attention to such marks (Gur *et al.* 2006; Nishikawa *et al.* 2006). Different approaches may be chosen to use correspondences found between views in a detection method. One way would be to add a multi view stage to the detection method in which new estimates of the confidence of CAD findings are made based on the single view classification and the obtained correspondence between views. Investigation of such methods is outside the scope of this paper.

5.2 Methods

In this section we will first give a short overview of our single view detection scheme. Sections 5.2.2 and 5.2.3 present the linking method, and Sec. 5.2.4 explains the evaluation procedure and the used data set. We end this section by briefly describing a technique to regroup suspicious regions marked by a CAD scheme (see Chapter 4). During the development of multiple view techniques we have found that these are hampered by the presence of multiple candidate regions on the same lesion or structure. The described regrouping technique removes many of these multiple regions.

5.2.1 Single view detection

Our CAD scheme consists of the following steps (see Fig. 5.1):

- Segmentation of the mammogram into breast area, pectoral muscle, and background area,
- Initial detection step resulting in a likelihood image $L_1(x, y)$
- Peak detection resulting in a number of suspect image locations,
- Region segmentation with dynamic programming using these locations as seed points,
- Final classification step to determine the likelihood of malignancy $L_2(i)$ for each region i.

These steps wills be described in more detail in the following paragraphs.

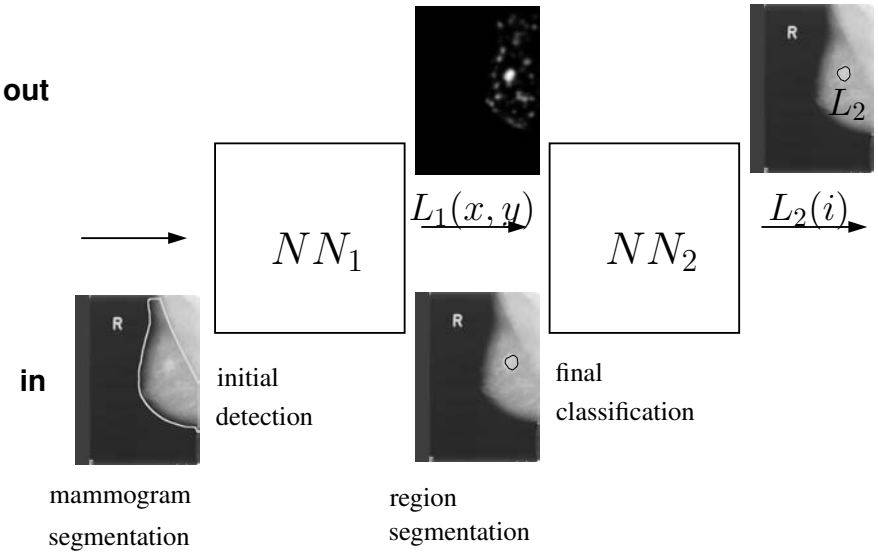


Figure 5.1: *Schematic overview of our single view CAD scheme. Here $L_1(x, y)$, the mass likelihood, is a measure of suspiciousness for every location in the breast area, which is obtained after application of a neural network classifier (NN_1). And $L_2(i)$, referred to as the likelihood of malignancy, is the final output of our single view detection scheme for each region, obtained after application of a second neural network classifier (NN_2).*

Segmentation of the mammogram The first step in our CAD scheme is the segmentation of each image into breast area and background, using a breast boundary segmentation algorithm that was developed previously in our group. In the MLO views, part of the pectoral muscle is visible, that needs to be segmented as well. To this end we use a pectoral muscle segmentation algorithm, which is described in detail in Karssemeijer (1998).

Initial detection step For each location inside the breast area a number of features are calculated that represent tumor characteristics. These features are related to the presence of spicules (Karssemeijer & te Brake 1996) and a central mass (te Brake & Karssemeijer 1999). A neural network (NN_1) is used to classify each location using these features and assigns a measure of suspiciousness to it, resulting in a so-called likelihood image $L_1(x, y)$. This likelihood image is smoothed, and a peak detection is performed at one scale. The peak detection algorithm visits every location in the likelihood image and a peak is detected when the likelihood is above a certain threshold and there are no other locations in the neighborhood (square neighborhood with size 6.4 x 6.4 mm) with a higher likelihood value. This results in a number of suspect image locations. An

algorithm searches for peaks that are located closer than 8 mm together, in order to remove multiple candidate locations on the same lesion. If these multiple regions are found the candidate location with the lowest likelihood value is removed.

Region segmentation The initial detection step results in a number of suspect image locations. Each of the detected peaks is used as seed point for region segmentation, based on dynamic programming (Timp & Karssemeijer 2004).

Final classification For each region, features are calculated that describe the position of a region in the breast (for instance the distance to the pectoral and the skin), region size, contrast, texture, compactness and acutance measures. A neural network (NN_2) is used to classify regions as true abnormalities and false positives based on these features. We use a soft classification in which a likelihood of malignancy (L_2) is determined.

5.2.2 Correspondence measures

We will describe the methods for the situation where the **MLO** view is studied and corresponding regions are sought in the **CC** view. In the other direction the procedure is exactly the same, and thus in the following **MLO** can be replaced for **CC** and vice versa.

For every region in the **MLO** view a search area is defined in the **CC** view based on the distance to the nipple. For every candidate region in the search area in the **CC** view, features are calculated that compare both regions (*joint features*). These features are used as input to a LDA classifier to discriminate between correct and incorrect links. Based on the LDA output, for every region in the **MLO** view the most likely link with a region in the search area of the **CC** view is sought. In the following paragraphs every step will be described in more detail.

definition of search area

Radiologists use the distance to the nipple to correlate a lesion in the MLO and CC view. It is generally believed that this distance remains fairly constant. Therefore, we use this distance to define an annular search area in the **CC** view. For all points within the search area in the **CC** view (see Fig. 5.2), the distance to the nipple is comparable to the distance between the candidate region and the nipple in the **MLO** view. To set the search area width, we used an annotated database containing 373 MLO/CC image pairs with a mass lesion that is visible in both views. For a lesion in the **MLO** view the corresponding lesion in the **CC** view is within the search area if the radial distances in both views do not deviate too much, i.e. the difference in radial distance to the nipple is less than half the search area width. Figure 5.3 presents the percentage of lesions in the **CC** view that is within the search area for varying width. Based on this, we set the width of the search

area for all cases to 48 mm. The nipple location was roughly estimated using a simple approach in which we assumed that the nipple is the point on the skin contour with the largest distance to the chest or the pectoral muscle (for the MLO views).

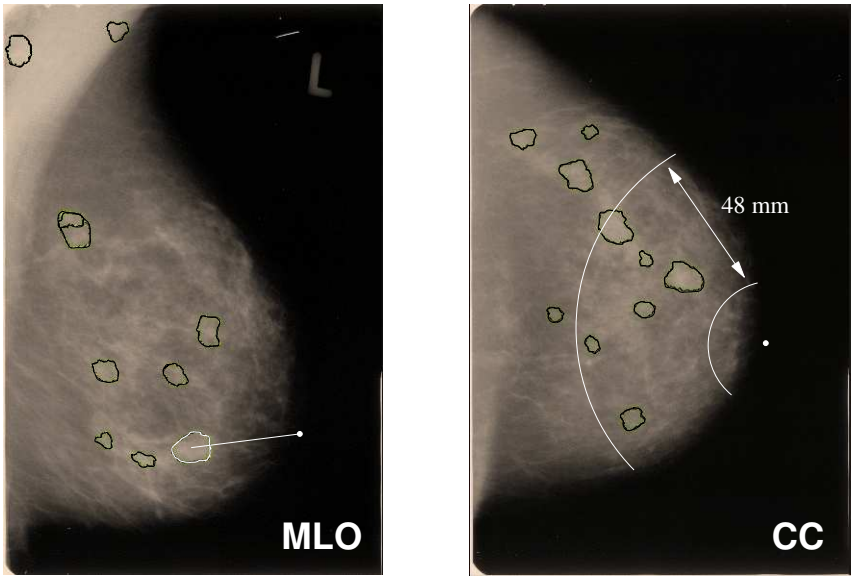


Figure 5.2: *Definition of the search area in the CC view. The estimated nipple location is indicated by the white dot in both views.*

calculation of joint features

Features for which the value in one view can be easily predicted from the value in the other view will be good features to establish correct links. Examples of such features are those that are invariant with respect to compression and positioning, and have high correlation between the values in both views. The following paragraphs describe the features that we investigated for this purpose.

*** distance to the nipple** Next to using the distance to the nipple for the definition of the search area, the distance to the nipple is also used as a joint feature. In a previous study, we have tested the assumption that the distance between a lesion and the nipple remains fairly constant between views (van Engeland & Karssemeijer 2001). We showed that there is a high correlation between the distance between the lesion and the nipple in MLO and CC views (correlation coefficient 0.90). The distance feature is defined as follows:

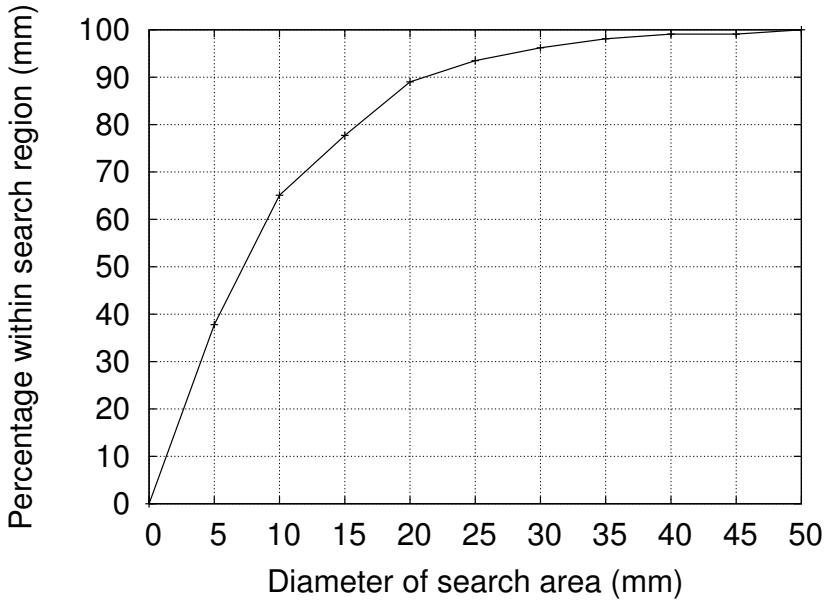


Figure 5.3: *Percentage within the annular search region for varying search area size.*

$$distance = \frac{|d_{mlo} - d_{cc}|}{w}, \tag{5.1}$$

where d_{mlo} is the distance between the region and the nipple in the MLO view, d_{cc} is the distance between the region and the nipple in the CC view, and w is the width of the search region. In Fig. 5.4 we have plotted the distance from a lesion to the nipple in the CC view against the distance in the MLO view for all lesions in the used data set.

*** angle** Although the correlation of the angular location in both views is much lower than that of the radial location (van Engeland *et al.* 2002), there is some correlation since the difference between the MLO and CC projection angle is 45 deg. This is shown in Fig. 5.4. Therefore, the angular location of a region is also tested as a joint feature (see Fig. 5.5). This feature is defined as follows:

$$angle = |a_{mlo} - a_{cc}|, \tag{5.2}$$

where a_{mlo} is the angular location of a region in the **MLO** view, and a_{cc} is the angular location of a region in the **CC** view.

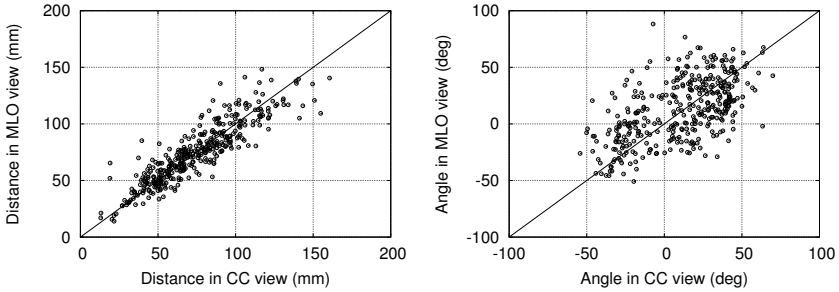


Figure 5.4: *(a) Distribution of the values of the distance feature in MLO and CC views. (b) Distribution of the values of the angle feature in MLO and CC views.*

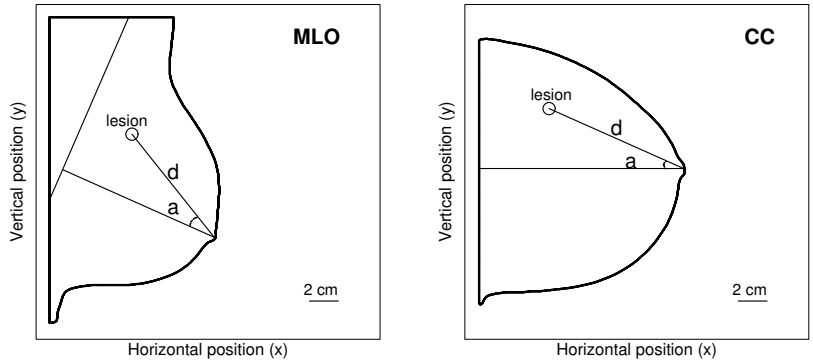


Figure 5.5: *Schematic representation of both mammographic views, indicating also the angular a and radial d location.*

*** gray scale correlation** Gray scale template matching was used by both Sanjay-Gopal *et al.* (1999) and Timp *et al.* (2005), who developed a regional registration technique for the registration of lesions in temporal pairs of mammograms. We implemented this registration measure as a feature for linking MLO and CC regions.

First a mass template is created which consists of all pixels enclosed by the contour of the region, and a band of pixels around the contour (Fig. 5.6). The width of this band

(a) (b)

Figure 5.6: *(a) Image showing the template used for gray scale template matching. (b) Image showing the inner and outer template used for histogram correlation.*

was equal to the maximum distance to the contour of all points inside the contour. We placed this template on a candidate region c in the **CC** view with center of mass of the **MLO** template on the peak in the likelihood image of the candidate region in the **CC** view, and calculated Pearson's correlation measure:

$$gray\_scale\_correlation = \frac{\sum_{x,y}(g_{mlo}(x,y) - \overline{g_{mlo}})(g_{cc}(x,y) - \overline{g_{cc}})}{\sqrt{(\sum_{x,y}(g_{mlo}(x,y) - \overline{g_{mlo}})^2)(\sum_{x,y}(g_{cc}(x,y) - \overline{g_{cc}})^2)}}, \tag{5.3}$$

where (x, y) is the location inside the **MLO** mass template, $g_{mlo}(x, y)$ is the pixel value in the **MLO** mass template at location (x, y), $g_{cc}(x, y)$ is the pixel value of the candidate region c at the same relative location. The average pixel values in the **MLO** mass template and the candidate region are given by $\overline{g_{mlo}}$ and $\overline{g_{cc}}$. In a recent study Filev *et al.* (2005) compared twelve different similarity measures for the task of template matching in temporal mammogram pairs. Pearson's correlation coefficient was one of the measures that performed best.

*** polar registration** The gray scale template matching described above has been used to find corresponding regions pairs in temporal mammograms. The problem of finding corresponding regions in MLO and CC views is different, however, since both views are acquired under compression at different angles. This, together with the fact that the

breast is a soft-tissue structure which is inhomogeneous, anisotropic and compressible, makes it very difficult to predict how the shape of the lesion on both projections will correlate. To compensate for this we have implemented a registration measure deduced from the standard template matching algorithm. Here the regions are transformed to a polar coordinate system with the center of mass of the regions as center. Then Pearson's correlation measure in the polar representation is calculated, allowing also a rotation ϕ of the **CC** region with respect to the **MLO** region. The resulting registration feature is described as:

$$\textit{polar\_correlation} = max(\textit{gray\_scale\_correlation}(\phi)). \tag{5.4}$$

In Fig. 5.7 an example is given of a candidate region in a **MLO** view and the corresponding candidate region in the **CC** view, and the gray scale correlation as a function of the rotation angle.

*** histogram correlation** To determine the histogram correlation between the region in the **MLO** view and the candidate region in the **CC** view, two templates are used, one containing the inside of the region and one containing a band of pixels outside the contour (see Fig. 5.6). To correct for differences in exposure and the local density of the backgrounds in which the lesions are embedded, for both the **MLO** mass template and the candidate region the cumulative distribution functions are obtained. From these cumulative distribution functions a look-up table of pixel values is determined that approximately maps pixel values of the **MLO** mass template to pixel values of the candidate region. Next, after application of the look-up table, the pixel value histograms for both regions are obtained. The histogram correlation (HC) is then calculated as:

$$HC = 1 - \frac{1}{2}\sum_{g} |\frac{H_{mlo}[g]}{T_{mlo}} - \frac{H_{cc}[g]}{T_{cc}}|, \tag{5.5}$$

where g is the pixel value, H_{mlo} is the histogram of the region in the **MLO** view, H_{cc} is the histogram of the candidate region in the **CC** view, and T_{mlo} and T_{cc} are the total number of counts in the **MLO** respectively the **CC** histogram.

The used joint feature is described as:

$$\textit{histogram\_correlation} = mean(HC_{in}, HC_{out}), \tag{5.6}$$

with HC_{in} the histogram correlation for the inner template, and HC_{out} the histogram correlation for the outer template.

*** mass likelihood** We refer to the output of the first neural network classifier in our detection scheme as the mass likelihood (L_1 in Fig. 5.1). From this mass likelihood

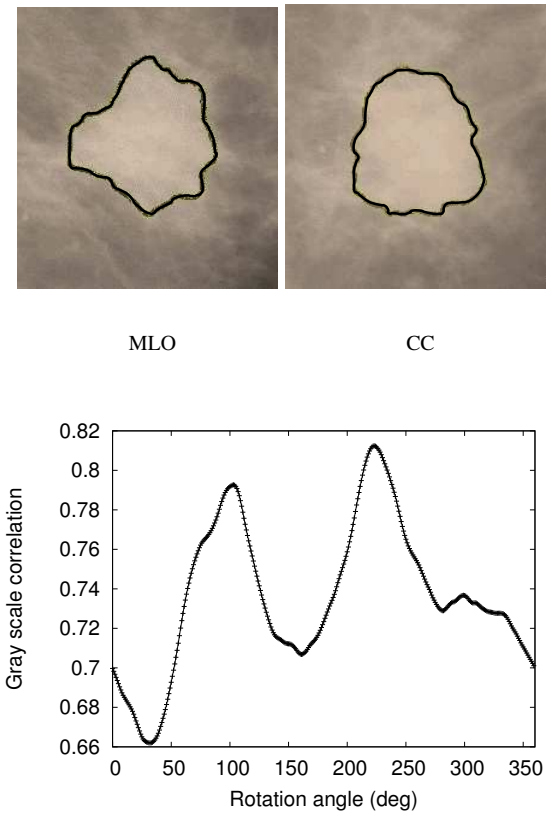


Figure 5.7: *An example of a region and the corresponding region in the other view, and the gray scale correlation as a function of the rotation angle of the candidate region in the CC view.*

two joint features are derived, the mass likelihood of the region in the **CC** view (*likelihood_other_view*), and the difference in mass likelihood of the regions in both views (*likelihood_difference*).

*** region size** The difference in size of the segmented regions in both views is also tested as a joint feature:

$$\textit{region\_size} = |RS_{mlo} - RS_{cc}|, \tag{5.7}$$

with RS_{mlo} and RS_{cc} the size of the segmented region in the **MLO** view, respectively the **CC** view. Since the **MLO** and the **CC** view are different projections of the same mass, we do not expect that the correlation between both sizes is high, however there might be a weak correlation.

*** contrast** Finally, the use of a contrast feature for linking is tested. The feature that describes the difference in contrast of the regions in both views is described as:

$$\textit{contrast\_difference} = |contrast_{mlo} - contrast_{cc}|, \tag{5.8}$$

with $contrast$ defined as

$$contrast = \frac{(E(I) - E(O))^2}{\sigma(I) + \sigma(O)}. \tag{5.9}$$

Here $E(I)$ and $E(O)$ are the average intensities inside and outside the contour, and the square of the difference between the average intensities are divided by the sum of the standard deviations. The outside region consists of all pixels outside the segmentation with a distance of less than $0.6R$ from the contour, where R is the effective radius ($R = \sqrt{\frac{\text{area of the inside region}}{\pi}}$) of the inside region.

5.2.3 classification of region combinations

The features described above are used as input into a LDA classifier to discriminate between correct and incorrect region pairs. A correct region pair is a true positive detection linked with a true positive (TP) detection in the other view. Incorrect region pairs are TP - false positive (FP) combinations. To select the best features for linking, forward feature selection (based on the area under the Receiver Operating Characteristic (ROC) curve, A_z) is used. To train the classifier only candidate region pairs are used where the region in the MLO and/or CC view is a lesion, this means only TP - TP, TP - FP and FP - TP region combinations. The classifier is tested using 50 % cross-validation. We refer to the resulting LDA classifier output as the *correspondence score*. Application of the LDA classifier thus results for every candidate region in the **MLO** view in a list of

possible links in the **CC** view with accompanying correspondence scores (see Fig. 5.8). We investigated the following three methods to obtain the final links of candidate regions from the correspondence score tables:

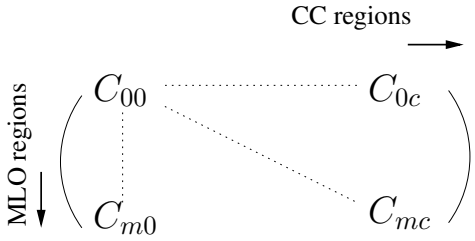


Figure 5.8: *Table of correspondence scores between regions in the MLO and CC view.*

I) determine most likely link for every candidate region For every detection in the **MLO** view the most likely link in the **CC** view is selected, that is the one with the highest correspondence score.

II) one-to-one correspondence This method assumes that a region can only be linked once to another region, so-called one-to-one correspondence. It starts with selecting the most likely combination of candidate regions from the correspondence score table, for instance between region c_0 from the **CC** view and region m_1 from the **MLO** view. For the next step all combinations with either c_0 or m_1 are excluded. The process continues until all **MLO** and/or **CC** regions are used.

III) one-to-one correspondence - maximize total correspondence score This methods also assumes one-to-one correspondence, however it searches for the combination of candidate regions so that the total of the correspondence scores selected is maximal. We used the Linear Assignment Problem (LAP) algorithm described by Jonker and Volgenant (Jonker & Volgenant 1987) for this purpose.

5.2.4 Performance testing

The initial detection step in our CAD scheme results in a number of suspect image locations. The number of image locations that is taken into account for further evaluation is determined by a threshold value for the likelihood. Here we used a relative low threshold, resulting in a relative large number of candidate regions, however, for each view we only take into account at most ten suspicious locations. We tested the effect of the number of candidate regions in the other view on the performance. The performance of the linking

algorithms are compared by counting the number of TP - TP links, and the number of TP - FP links established.

The used data set consisted of 412 cases, containing 1648 images. All cases had four-view mammograms. In every case there was a mass lesion visible in at least one of the views. The mammograms were taken from two annotated databases containing cases from the Dutch breast cancer screening program. The annotations were made by or under supervision of an expert radiologist. Both databases contained mammograms obtained at three points in time, the *diagnostic* mammogram at time of detection and the mammograms taken in the two screening rounds prior to detection (referred to as *prior* and *reference* mammogram). The diagnostic mammograms were either clinical mammograms of the interval cancer cases or screening mammograms of the screen detected cases. The time interval between subsequent screening mammograms of each case is two years on average. In this study, the majority of the mammograms used were diagnostic mammograms, and a small percentage were prior mammograms.

The mammograms from the first database were digitized with a Lumisys 85 digitizer at a pixel resolution of 50 μm, and averaged down to a resolution of 200 μm, maintaining the original gray value depth of 12 bits. The mammograms from the second database were digitized with a Canon CFS300 scanner with the same resolution and gray value depth.

Table 5.1: *The used data set.*

set	no. images	no. cases
database 1 diagnostic	740	185
database 1 prior	112	28
database 2 diagnostic	744	186
database 2 prior	52	13
total	1648	412

5.2.5 Removing multiple candidate regions on the same lesion

During the development of multiple view techniques we have noticed that they are hampered by the phenomena that mass lesions are sometimes detected by multiple candidate regions. In the previous chapter we have developed a technique to regroup initial CAD locations to facilitate the final classification of suspicious regions (van Engeland & Karssemeijer 2005). This regrouping technique searches for candidate locations that belong to the same structure. Therefore, it takes into account the distance between the

locations and the image structure along a path between the locations. When correspondence is found, the two locations are replaced by a new location in between the initial candidate location. In this paper we investigate further the effect of using this algorithm prior to linking.

5.3 Results

Table 5.2: *Link results of the LDA classifier for five candidate regions in the other view. The features that were selected by forward feature selection were: likelihood_other_view, polar_correlation, distance, histogram_correlation, and contrast_difference.*

features	A_z	number TP - TP links
distance	0.743	481 (55 %)
angle	0.510	261 (30 %)
gray_scale_correlation	0.725	562 (64 %)
histogram_correlation	0.606	541 (62 %)
polar_correlation	0.782	601 (69 %)
likelihood_other_view	0.826	681 (78 %)
likelihood_difference	0.609	456 (52 %)
region_size	0.510	249 (28 %)
contrast_difference	0.501	256 (29 %)

We used forward feature selection to find the optimal set of joint features. To give an impression of the importance of the different features for the linking method, Table 5.2 presents the individual performances. It can be seen that the mass likelihood of the candidate region in the other view is a good indicator for the correctness of a link. This was expected since the classifier was trained on TP - TP and TP - FP combinations, and the TP candidate regions are more likely to have a higher mass likelihood value. The distance to the nipple was also an important joint feature. An interesting result that follows from Table 5.2, is that the polar gray scale correlation measure performs also very good, showing a higher A_z than the standard gray scale correlation. Region size and contrast appeared to be less important correlation features. The forward feature selection procedure resulted in the selection of five joint features, namely *likelihood_other_view*, *polar_correlation*, *distance*, *histogram_correlation*, and *contrast_difference*. Using these features in the LDA classifier the A_z value for the link classifier was 0.92.

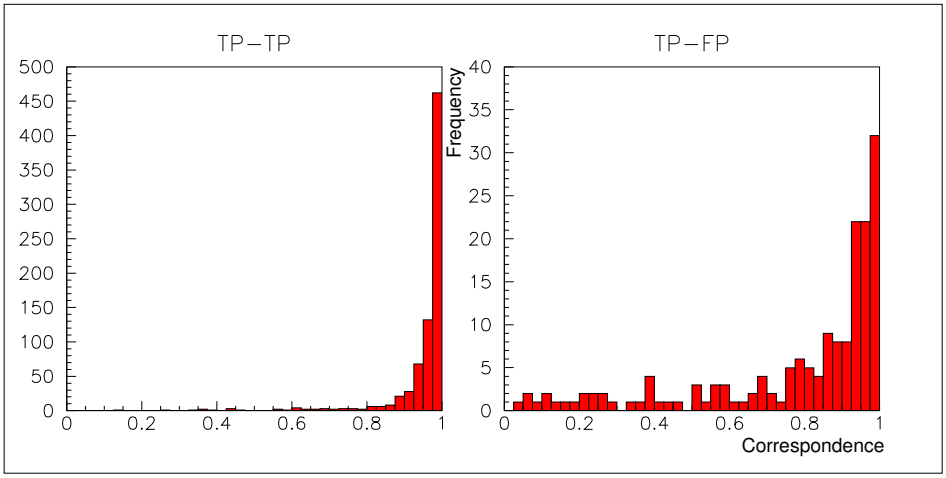


Figure 5.9: *Distributions of the correspondence score for two types of region combinations.*

After application of the LDA linking scheme using the selected features, for every candidate region in the **MLO** view a list of possible links in the **CC** view with accompanying correspondence scores was obtained. Figure 5.9 presents the outputs of the LDA classifier for two types of region combinations.

Table 5.3: *Link results for different combination methods. The second half of the table presents the link results after application of the regrouping technique described in Sec. 5.2.5.*

method	TP - TP	TP - FP
most likely link for every candidate region	707 (81%)	162 (19%)
one-to-one correspondence	531 (61%)	335 (39%)
one-to-one correspondence - maximize total score	415 (47%)	451 (53%)
after regrouping technique		
most likely link for every candidate region	699 (82%)	155 (18%)
one-to-one correspondence	563 (66%)	289 (34%)
one-to-one correspondence - maximize total score	406 (47%)	444 (53%)

We tested three methods to obtain the final region links from the correspondence score tables. Table 5.3 presents the results for the three methods. It can be seen that selecting for every region in the **MLO** image the most likely link in the **CC** image works

better than the other two methods, which require a one-to-one correspondence between the regions in the two views. 82 % of the TP detections were correctly linked to the TP detection in the other view. In total there were 872 TP regions in the data set. For some cases there was no TP detection in one of the views, for instance, when the lesion was located near the pectoral muscle in the MLO view. Since the pectoral muscle is often not visible in CC views, it occurs often that these lesions are not visible in the CC view. On the other hand, for other cases there was more than one TP detection in the MLO and/or CC view. Because not all cases contained a TP region in both views, the maximum performance that could have been achieved is 92 %. Table 5.3 also presents the results of linking after application of the technique to regroup the initial candidate regions. For the data set used, regrouping only gives gives a very small increase in the link performance. When looking at the number of cases with incorrect links (TP - FP combinations), the effect of the regrouping technique is more clear. Before regrouping there where 137 (out of 412) cases containing a TP - FP combination, and after regrouping this number reduced to 129 cases.

We also tested the effect of the number of candidate regions in the other view. Figure 5.10 shows that the fraction of correct links increases with the number of candidate regions in the other view until five regions per view, for more than five candidate regions the performance stays more or less constant.

We investigated the causes for missed links, and present the results in Table 5.4. When a rather subtile lesion was incorrectly linked to a FP region in the other view we counted this as the first error cause presented in the table. This was because the lesion appearance was in these cases actually not very different from the appearance of a glandular tissue part. The fourth cause of errors mentioned, were the cases were the lesion was clearly visible in both views for an observer, however the appearance was very different on both views. In Fig. 5.11 an example of a MLO/CC image pair is given where the true positive detections in the CC view are incorrectly linked to regions in the MLO view. For this case the incorrect links are mainly due to the large difference in appearance of the lesion in both views. We also found that for the prior mammograms the percentage of correct links was lower than for the diagnostic cases (69 % correct against 83 % correct). The lesions in these mammograms are less obvious since they were detected by the radiologists in retrospect.

5.4 Discussion & conclusions

We have presented a method to link candidate regions determined by our CAD scheme in MLO and CC projections. This method uses a LDA classifier to discriminate between correct and incorrect links, based on a set of link features. This classifier results for every possible combination of candidate regions in a *correspondence score*. Using an

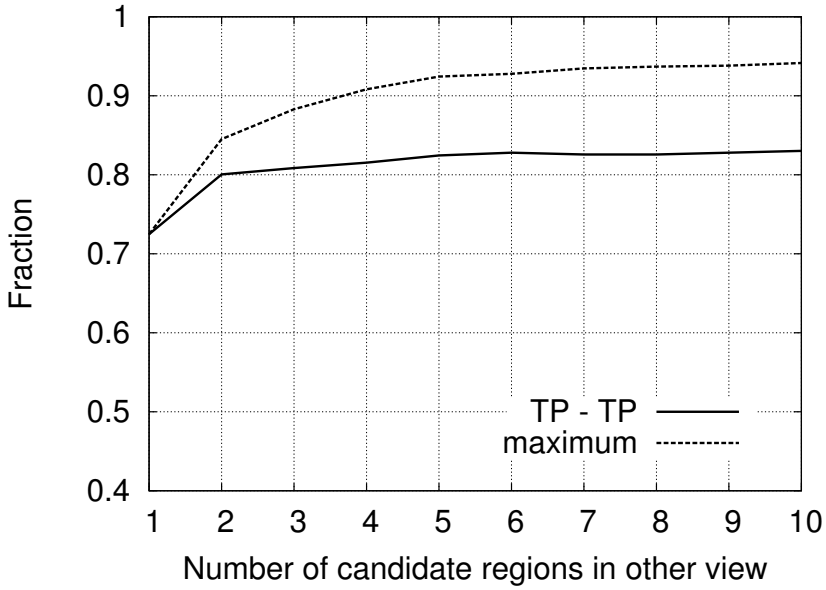


Figure 5.10: *The fraction of correct and incorrect links for varying number of regions in the other view. Here we tested for every view containing a lesion whether this lesion was correctly linked with the lesion in the corresponding view. The dotted line indicates the fraction of links that would have been found if all detected true positives were correctly linked.*

annotated database we have shown that the method was able to establish a correct link between the true positive regions for 82 % of the TP regions with the TP region in the other view.

In Sec. 5.2.3 we have described three methods to obtain the final links of the candidate regions from the correspondence score tables. We found that allowing only one-to-one correspondences (both methods II and III) between candidate regions in corresponding view decreases the number of correct TP-TP links considerably (see Table 5.3). This may partly be explained by problems related to lesions marked by more than one candidate region. The main disadvantage of one-to-one correspondence, however, lies in the fact that sometimes the correspondence of a true positive with a false positive is stronger than the correspondence with the actual true positive. By just using the most likely link for every region the true positives may still be linked, whereas using one-to-one correspondence this is not allowed. Using the LAP algorithm appears to have worse performance. This suggests that maximizing the total correspondence score is not a good

Table 5.4: *The most important causes for link errors.*

similar region at same distance from nipple	33 %
large distance to nipple difference	33 %
multiple regions on the same lesion	5 %
lesion appearance very different in both views	3 %

idea because generally false positives in MLO and CC views do not correspond.

The majority of the incorrect links were caused by the fact that there was a similar region at about the same distance from the nipple and that the distance between the lesion and the nipple differed too much between both views (see Table 5.4). The latter could, for instance, be due to suboptimal positioning of the breast. Other reasons were multiple regions on the same lesion, and a strong difference in appearance between views. To reduce the effect of multiple regions per lesion we investigated application of a method that we developed previously to regroup initial candidate regions. The effect of this method is small on the data set used in the current study. This might be due to the fact that we used a data set that mainly contained small lesions, which are less likely to be detected by multiple candidate locations. The percentage of lesions that was prompted more than once by our CAD system is twice as low in the current study (10 percent against 20 percent in the referred study (van Engeland & Karssemeijer 2005)).

Even if lesions appear rather similar visually this may be difficult to determine for an automatic method. In particular if features are used that strongly rely on lesion segmentation similarity may be hard to establish, because segmentation of lesions is a notoriously complex problem. In this work we tried to avoid the use of features that are strongly dependent on region boundaries. Furthermore, we by using a state-of-art segmentation method (Timp & Karssemeijer 2004) we tried to eliminate variability due to inaccurate segmentation as much as possible.

One of the features that was selected for the link classifier was the difference in distance between the region and the nipple in both views. To calculate this distance we used automatically determined nipple locations, using a simple algorithm in which we assumed that the nipple is the point on the skin contour with the largest distance to the chest or pectoral muscle. To determine the effect of the automated nipple detection method on the link results, for the subset *database 1 diagnostic* we also determined the number of correct links when using manually indicated nipple locations. For this subset the percentage of correct links was 85 % with the manually determined nipple locations and 84 % with the automatically determined nipple locations. Based on these results we assume that the nipple detection method had little influence on our final link results.

It is very difficult to predict the shape of a lesion in the CC view from the shape of

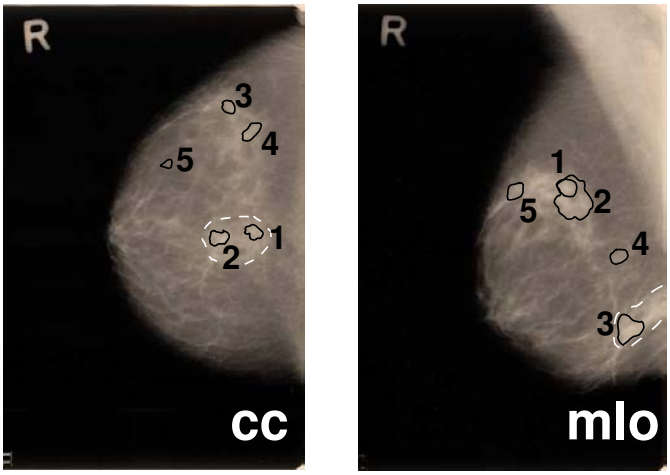


Figure 5.11: *Example of a MLO/CC image pair showing the single view candidate regions. For this case our algorithm was not able to establish correct links between the TP detections in the MLO view and CC view. The white dashed line represents the annotation of the lesion made by the radiologist. The link results of our algorithm were:*
1 (in CC) ↔ 1 (in MLO), 2 ↔ 3, 3 ↔ 1, 4 ↔ 2, 5 ↔ 2.

the lesion in the MLO view. However, if lesions are somewhat elongated or irregular, this is usually seen in CC and MLO views, but orientation is different in general. Therefore we decided to implemented a template matching algorithm in which we also allow rotation. The resulting polar gray scale correlation measure proved to be a good feature to discriminate between correct and incorrect region combinations (see Table 5.2).

There are two studies in literature describing the correlation of MLO and CC views. Highnam *et al.* (1998) used a model-based method to find the curve in the MLO view on which a location in the CC view may be mapped. They demonstrated with a data set of 32 cases that the average minimum distance between this curve and the actual corresponding position was 6.48 mm. However, assumptions on the parameters and the deformation of a compressed breast had to be made, and to our knowledge further evaluation of the model has not yet been presented. Good *et al.* (1999) reported an attempt of matching computer-detected objects in two views. Their study was similar to ours. They also demonstrated the feasibility of identifying corresponding objects (A_z = 0.82) in two views by exhaustive pairing of the detected objects and feature classification using a database of 60 MLO/CC pairs containing 38 masses. The features they used for linking were related to location, contrast, region size and boundary complexity. Our ROC evaluation resulted in a higher A_z value of 0.92. Results suggest that the improved performance

is due to the use of the region likelihood and the gray scale correlation measures. When removing the region likelihood feature and the gray scale correlation features from our selection we found that the number of correct links decreased with 10 % and 3 %.

A preliminary investigation indicates that it is possible to improve mass detection results of our CAD scheme using the described method to link candidate regions in MLO and CC views. An application of the method may be found in improvement of the presentation of CAD results on a mammographic workstation. Studies suggest that radiologists are less likely to ignore prompts on true positives when CAD marks a lesion in both views (Gur *et al.* 2006; Nishikawa *et al.* 2006). The proposed method cannot be used directly, however, to increase lesions prompted in two views without increasing the number of false positives. To overcome this limitation more elaborate techniques are needed to combine information from two views in a CAD scheme. The method presented here to establish correspondence between regions should be regarded as one of the key component of such a scheme.

Chapter 6

Using information from two mammographic views to improve mass detection results [1]

6.1 Introduction

In the previous chapter we presented a method to link potentially suspicious areas determined by our CAD scheme in mediolateral oblique (MLO) and craniocaudal (CC) views. Joint features for all possible combinations of candidate regions in both views were calculated. These features described the difference in radial distance from the candidate regions to the nipple, gray scale correlation between both regions, and the mass likelihood of the regions. Next, a Linear Discriminant Analysis (LDA) classifier was used to give a correspondence measure for every possible combination. Finally, for every region in the original view, the region in the other view with the highest correspondence score was selected as the corresponding candidate region. Using an annotated database we have shown that this linking method was able to correctly link 82 % of the true positive (TP) regions in MLO and CC views.

In this chapter we describe a method that uses the obtained correspondences to improve the computer-aided detection of masses. In literature some approaches have been described to establish correspondence between multiple views. However, few studies have been described that use this correspondence to improve detection or classification results. Paquerault *et al.* (2002) developed a two view matching method which results in a correspondence score for each possible mass pair. By combining this correspondence

[1] The content of this chapter has been submitted to Medical Physics.

score with their single view detection score, their detection results improved significantly. Hadjiiski *et al.* (2001b) developed a method for the interval change analysis of masses to improve their classification accuracy between benign and malignant. Recently, Timp & Karssemeijer (2006) presented a study on interval change analysis. They used a regional registration technique and calculated temporal features to improve the detection performance.

To combine information from MLO and CC views, we investigate adding another classifier to our CAD scheme. As input to this classifier features that depend on combined information from MLO and CC views are used. The final detection performance is compared with our single view detection performance using Free-response Receiver Operating Characteristic (FROC) analysis.

6.2 Methods

In the previous chapter we described our single view detection scheme (see Sec. 5.2.1). To this scheme we added another classifier, which we will refer to as the *two view classifier* (NN_3 in Fig. 6.1). As input to the this classifier we have investigated both single view features and two view features. The two view features are the likelihood of malignancy of the corresponding region in the other view, and a number of measures that describe the resemblance between both regions. We will describe the used features in Sec. 6.2.1, and the two view classifier in Sec. 6.2.2. Finally, Sec. 6.2.3 describes the evaluation procedure and the used data set.

6.2.1 Features for two view classifier

As input to the two view classifier fifteen features are investigated (see Table 6.1). The two view features were described in detail in the previous chapter (see Sec. 5.2.2 and 5.2.3). We will give a short description of the used single view features, which originate from the first and second step of our CAD scheme, in the following paragraphs.

*** spiculation features** With respect to the detection of mass lesions in mammograms there are two important lesion characteristics, one is the presence of spicules and the other is the presence of a central mass. Our spiculation features are based on the idea that stellate lesions show a pattern of lines directed towards the center of a lesion. In the first step of our CAD scheme we use two spiculation features. The first feature is a normalized measure for the fraction of image locations with a line orientation directed towards the center. The second feature calculates to what extent the locations with a line orientation towards the center are equally distributed in all directions. We will refer to

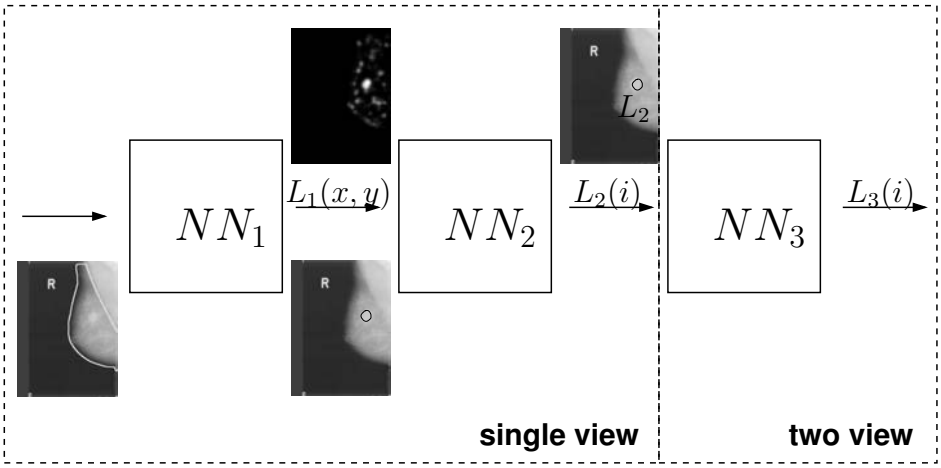


Figure 6.1: *Schematic overview of our CAD scheme. Here $L_1(x, y)$ is a measure of suspiciousness for every location in the breast area (mass likelihood). $L_2(i)$ is the final output of our single view detection scheme for each region (likelihood of malignancy). And $L_3(i)$ is the final output of our two view detection scheme.*

these features as f_1, respectively f_2. Details can be found in Karssemeijer & te Brake (1996).

*** focal mass features** For the detection of masses we use a similar approach as for the detection of spicules. Instead of determining the line orientations, we now calculate the gradient orientation at each location in the image. If a mass is present, the majority of image locations in a neighborhood of the center of the mass will have a gradient orientation towards the center. We derive two features from the calculated gradient orientations. The first feature g_1 is a normalized measure of the fraction of image locations with an intensity gradient pointing towards the center. The second feature g_2 calculates whether these locations occur in all directions of the center location. Details can be found in te Brake & Karssemeijer (1999).

*** mass likelihood** In the first step of our CAD scheme, the above described spiculation and mass features are used as input to a 3-layer backpropagation neural network (NN_1 in Fig. 6.1) trained on known abnormalities, as was described in Sec. 5.2.1 of the previous chapter. Application of this classifier results in a *mass likelihood* measure (L_1 in Fig. 5.1), which assigns a high value to suspicious locations.

Table 6.1: *The features we investigated for the two view classifier. The first column represents the features that originate from our original single view CAD scheme, and in the second column two view features are presented. The correspondence score is the output of the LDA classifier that was used to discriminate between correct and incorrect region combinations.*

single view features	two view features
- spiculation features (f_1,f_2)	- $correspondence$
- focal mass features (g_1,g_2)	- difference in distance to nipple ($distance$)
- mass likelihood (L_1)	- $gray\_scale\_correlation$
- contrast measures	- $polar\_correlation$
($contrast_1$, $contrast_2$)	- $histogram\_correlation$
- $region\_size$	- likelihood of malignancy single view of
- likelihood of malignancy	corresponding region ($L_2\_other\_view$)
single view (L_2)	

*** region size** Another feature that is used in the second step of our CAD scheme is the $region\_size$. This is the size of the segmented region, obtained after segmentation using dynamic programming (Timp & Karssemeijer 2004).

*** contrast measures** In the second step of our single view CAD scheme a number of contrast features are used. For the two view classifier two contrast measures are investigated. The first contrast measure is the difference in mean pixel value (which is linear with the optical density) between pixels inside ($E(I)$) and pixels in an area outside ($E(O)$) the segmentation,

$$contrast_1 = E(I) - E(O). \tag{6.1}$$

The area outside the segmentation consists of all pixels with a distance of less than $0.6R$ from the contour, where R is the effective radius ($R = \sqrt{\frac{\text{area of the inside region}}{\pi}}$) of the region. The second contrast feature is the square of the difference between the mean pixel value inside and outside the contour, divided by the sum of the standard deviations of both areas,

$$contrast_2 = \frac{(E(I) - E(O))^2}{\sigma(I) + \sigma(O)}. \tag{6.2}$$

*** single view likelihood of malignancy** Next to the classifier output of the first step, L_1, also the final output of our single view CAD scheme, L_2, is investigated as input to the two view classifier.

6.2.2 Two view classifier

Application of the linking algorithm results for every region in a corresponding region in the other view with accompanying correspondence score. If there is no corresponding link, the correspondence score and other two view features are set to zero.

To select the features to be used in our two view classifier, a Receiver Operating Characteristic (ROC) analysis is performed. For this purpose a LDA classifier is used. The optimal set of features is selected, using 50 % cross-validation, based on the area under the ROC curve (A_z) using a forward selection algorithm. To this end, a database of annotated cases is used, which is described in the following section.

The two view classifier is a 3-layer neural network (NN_3, in Fig. 6.1), trained with backpropagation and three hidden nodes. The input to this classifier are the features selected by the ROC analysis. Just as for the single view scheme, the output of this classifier is a likelihood of malignancy for every region.

6.2.3 Evaluation

Our two view detection scheme is evaluated on a data set containing 412 abnormal cases, and 537 normal cases. The abnormal cases are the same cases that were used in the previous chapter for evaluation of the linking algorithm (see Table 5.1). Again all cases have four-view mammograms. The set of normal cases were taken from the same two databases of mammograms that were used for the selection of abnormal cases. The normal cases do not include benign lesions, and were all verified to be normal by an experienced radiologist.

Table 6.2: *The data set that was used to test the two view classifier.*

set	no. images	no. cases
database 1 normal	140	35
database 2 normal	2008	502
normals	2148	537
database 1 diagnostic	740	185
database 1 prior	112	28
database 2 diagnostic	744	186
database 2 prior	52	13
abnormals	1648	412

The two view classifier is tested using crossvalidation with 95 percent training and

five percent testing. The performance of the two view classifier is compared with our original detection performance using FROC analysis and we present both an image and a case based evaluation. In the case based evaluation, a case is by definition regarded as a true positive (TP) case if in at least one of the two views the lesion is detected by our CAD scheme.

Both the output of our single view and our two view CAD scheme are standardized using only the images from the normal cases. To this end, we compute for every region the number of normal regions per image with values lower than that of the current region. We refer to this as the $normality\ score$. So, in other words, this is the frequency of occurrence in normal mammograms of regions that are at least as suspicious as the region at hand.

In the previous chapter we described three ways to obtain the final combinations of candidate regions from the correspondence score tables (Sec. 5.2.3). For a maximum of five regions per view, the effect of these linking methods (Sec. 5.2.2) on the two view detection results is also investigated.

Finally, we prepared a data set which contained only normals and cases with correctly linked TP regions (where the regrouping technique was used before linking). Also for this data set the performance of the two view classifier and the original single view detection performance are compared.

6.3 Results

Using forward feature selection, we determined the optimal set of features for our two view classifier. To give an impression of the importance of the different features for use in the two view classifier, Table 6.3 presents the individual feature performances. The output of our single view CAD scheme L_2 was the first feature that was selected by the feature selection algorithm. The features that were selected next were all two view features, namely $correspondence$, $polar\_correlation$, $L_2\_other\_view$, $distance$, and $histogram\_correlation$.

We compared the performance of the two view classifier with our original CAD performance using FROC analysis. The result of the image based evaluation is presented in Fig. 6.2, and shows an improvement when using the two view classifier. For instance, at a false positive rate of 0.1 FP/image the image based sensitivity increases from 52 % to 61 %. In the case based evaluation, however, we found no improvement.

To present the effect of the two view classifier in another way, in Fig. 6.3 and 6.4 the normalized classifier outputs of the single and the two view classifier are compared for both TP and FP regions. In the histogram for the TP regions (Fig. 6.3), a shift of the normality scores to the left can be seen. The negative values in the histogram correspond to TP regions where the normality score was decreased by the two view classifier, i.e.

Table 6.3: *A_z values for the LDA classifier for the single and two view features. Using forward feature selection the following features were selected: L_2, correspondence, $polar\_correlation$, $L_2\_other\_view$, distance, and $histogram\_correlation$.*

features	A_z
single view features	
f_1	0.547
f_2	0.594
g_1	0.753
g_2	0.774
L_1	0.837
$region\_size$	0.737
$contrast_1$	0.795
$contrast_2$	0.750
L_2	0.899
two view features	
$L_2\_other\_view$	0.765
$distance$	0.619
$gray\_scale\_correlation$	0.676
$histogram\_correlation$	0.626
$polar\_correlation$	0.742
$correspondence$	0.826

they have become more suspicious. This is consistent with the improved FROC curve presented in Fig. 6.2. In the case based evaluation, for every case the minimum normality score of the TP regions in the MLO and CC view was used. Here the shift to the left is no longer visible. Figure 6.4 shows the effect of the two view classifier on the FP regions, where the FP regions are sorted into four groups based on the original (single view) normality score. For the relatively suspect false positives (normality score > 0.1 & < 1.0) the histogram shows a shift to the right (they are found less suspicious by the classifier). However, part of the less suspicious regions (normality score > 1.0) are upgraded by the two view classifier.

In Figure 6.5 the single view detection results are compared with the two view de-

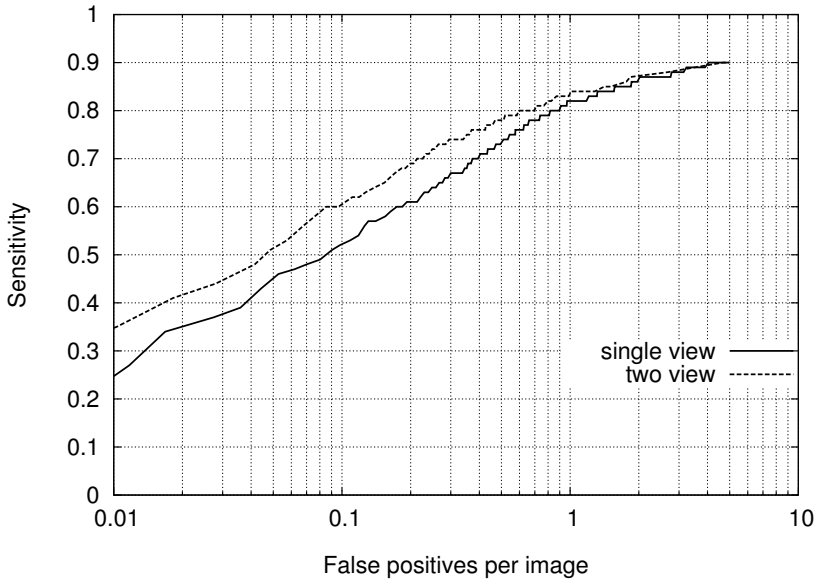


Figure 6.2: *Single view and two view mass detection results (image based FROC evaluation). The following features were used as input: L_2, correspondence, $polar\_correlation$, $L_2\_other\_view$, distance, and $histogram\_correlation$.*

tection results for different linking methods. It is shown that the improvement of the detection results by the two view classifier is highest when using linking method I. In Sec. 5.2.3 it was presented that this method, where for each region combinations with all candidate regions in the other view were possible, performed best. The number of correctly established TP - TP regions links was 25 % higher than for the other two methods, which assumed one-to-one correspondence.

Finally, Fig. 6.6 presents the results of the FROC analysis when using the part of the data set with cases in which the TP regions were correctly linked. Here, both the image and case based evaluation show an improvement by application of the two view classifier.

6.4 Discussion & conclusions

We have presented a method to combine information from MLO and CC views in our CAD scheme to improve detection results. A two view classifier was added to our single view CAD scheme. The input to this two view classifier was a feature vector containing the likelihood of malignancy of the region, the likelihood of malignancy of the corresponding region in the other view, and a number of features that describe the resemblance

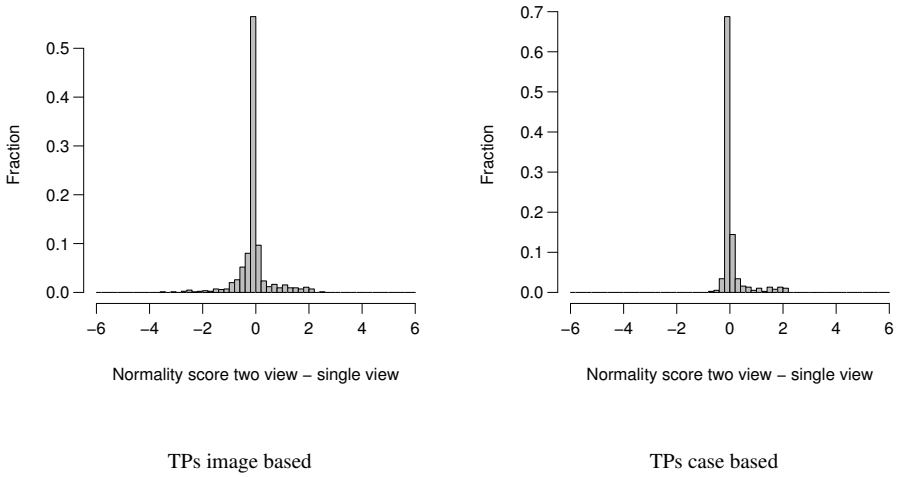
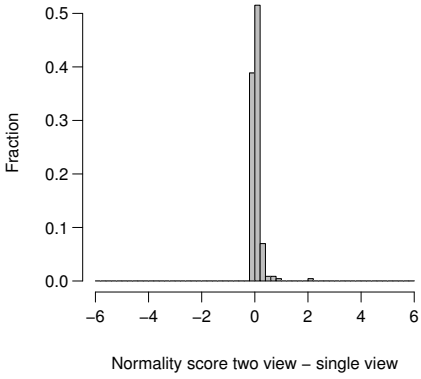


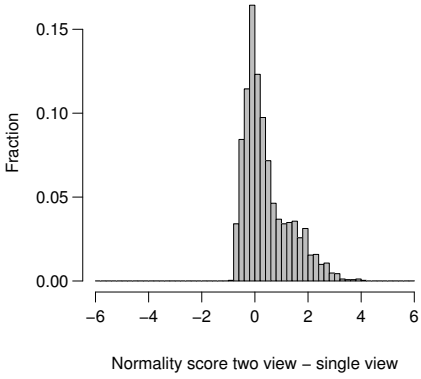
Figure 6.3: *Histograms of difference between normality scores before and after application of the two view classifier for true positives per image and per case.*

between the both regions. In order to obtain region correspondences between regions in MLO and CC views we used an algorithm that was described in the previous chapter. The output of the two view classifier was a likelihood of malignancy for every region. Using FROC analysis, we showed that our detection results improve for the image based FROC evaluation when using two view information (see Fig. 6.2).

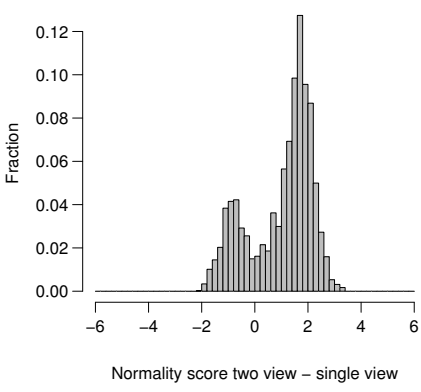
In the case based evaluation we found no improvement. It might be that lesions that were found more suspicious after application of the two view classifier were already very suspicious in the other view. Consequently, in the case based evaluation the sensitivity for a certain false positive level did not change. Therefore, this might be an explanation of the fact that the case based performance did not change. As can be seen in Fig. 6.3 the positive effect of the two view classifier on the malignancy score of the TP regions is not visible in the case based evaluation. On the contrary, we see a small negative effect. Figure 6.4 shows that the effect of the two view classifier on the FP regions also has a positive effect on the performance. However, it seems that this benefit is canceled by the negative effect on the TP regions in the case based evaluation. This effect can be seen more clearly when calculating the sum of differences of the normality scores of the TP regions before and after application of the two view classifier. In the image based evaluation this sum of differences is 38.4 against -38.9 in the case based evaluation.



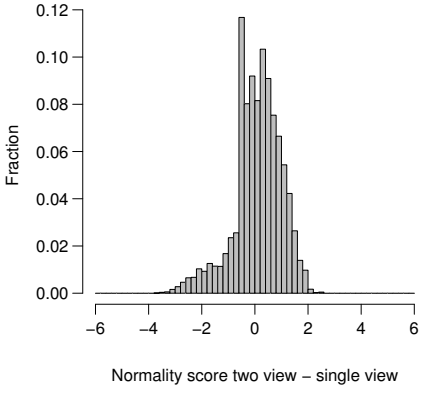
FPs, normality score < 0.1



FPs, normality score > 0.1 & < 1.0



FPs, normality score > 1.0 & < 2.5



FPs, normality score > 2.5 & < 5.0

Figure 6.4: *Histograms of difference between normality scores before and after application of the two view classifier, for different groups of false positives.*

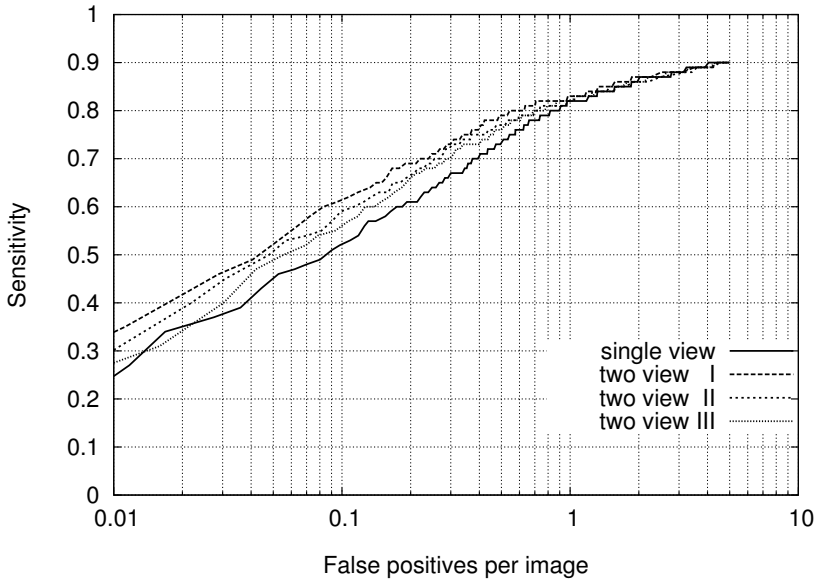


Figure 6.5: *Image based evaluation for single view and two view detection using different linking methods, with I,II and III referring to the linking method described in Sec. 5.2.3.*

For the two view classifier it is important that the number of incorrect links between a TP region in one view and a FP region in the other view is as low as possible. This can be seen in Fig. 6.5, where we compared the different linking methods. Here the use of the linking method with the highest percentage of correctly established TP - TP region links gave the highest increase in detection performance. This can also be seen in Fig. 6.6, where we present the results of the two view classifier on the subset of cases with correct TP - TP region links. Here also the mass detection results in the case based evaluation improved.

To our knowledge, the only study presented so far on the use of MLO and CC information to improve detection results is that of Paquerault *et al.* (2002). They developed a two view matching method which results in a correspondence score for each possible mass pair. By combining this correspondence score with their single view detection score, their classification results also improved significantly. Their image based detection sensitivity was found to improve from 62 % with a one view detection scheme to 73 % with their two view scheme, at a false positive rate of 1 FP/image. The corresponding case based detection sensitivity improved from 77 % to 91 %. These results seem to be in contradiction with our results. Figure 6.2 shows that at the false positive level of 1 FP/image the sensitivity of our scheme is improved only very slightly by using two view

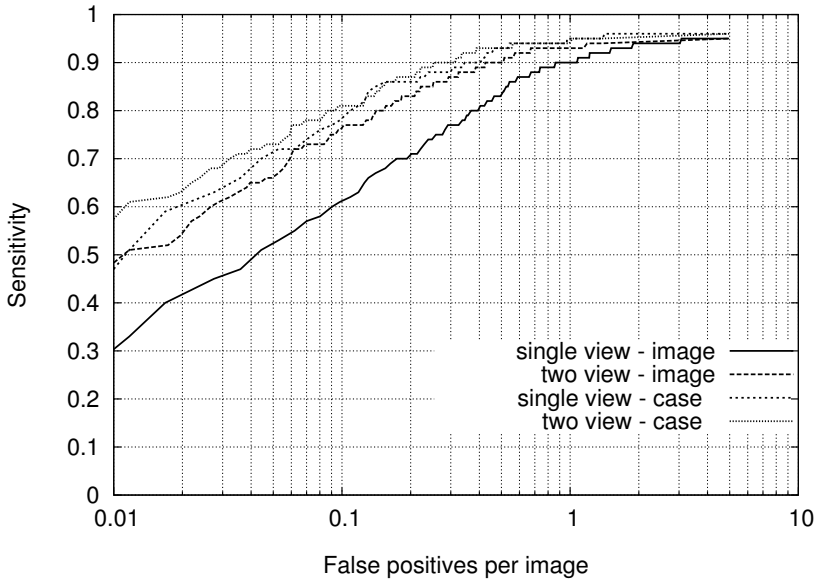


Figure 6.6: *Image based evaluation for single view CAD and two view after removing incorrect links.*

information (from 82 % to 83 %). This requires further investigation. Unfortunately, in the paper by Paquerault *et al.* no information was given about the percentage of correctly established links.

Our detection results mainly improved for operating points with high specificity (low number of false positives per image) in the FROC curve. For instance at a false positive rate of 0.1 FP/image, the image based sensitivity increased from 52 % to 61 %. Since it is important to obtain an as high as possible sensitivity without giving too many false positive prompts to the radiologist, this is also the part of the curve we are most interested in an improvement of the detection.

To summarize, our two view mass detection method improved the detection results for the image based FROC evaluation. We have found that it is harder to improve the results in a case based evaluation. Our method leads to more consistent CAD malignancy scores for MLO and CC views, especially in the first part of the curve. When presenting CAD results to radiologists they tend to complain when a lesion is marked by the CAD system in one view and not in the other, although the lesion appears rather similar visually in both views. The improvement we found in the image based FROC curve leads to a more consistent behavior of the CAD system.

Chapter 7

Using context for mass detection and classification in mammograms [1]

7.1 Introduction

In mammography, CAD mass detection and classification techniques mainly use local image information to determine whether a region is abnormal or not. There is a lot of interest in developing CAD methods that use image context, asymmetry, and multiple view information. Radiologists in breast cancer screening are trained to use comparisons of the left and right breast to identify suspicious asymmetric densities. It is also known that screening with two mammographic views, mediolateral oblique (MLO) and craniocaudal (CC), improves the detection accuracy of abnormalities in the breast, which can be explained by the fact that two projections allow better estimation of conspicuity of lesions and may reveal lesions hidden by glandular tissue in one of the projections. It is expected that human readers take advantage of their ability to easily match lesions in different views. It is common for CAD algorithms to combine detection results in separate views independently.

A number of studies have been done to investigate the use of context for CAD, including asymmetry and multiple view information. Paquerault *et al.* (2002) investigated the use of two-view information, combining MLO and CC, to improve CAD mass detection. They developed a two view matching method which resulted in a correspondence score for each possible mass pair. By combining this correspondence score with their

[1] The content of this chapter has been published previously in van Engeland *et al.* (2005).

single view detection score, their results improved significantly. At a false positive rate of 1 FP/image they found an increase in sensitivity from 62 % to 73 %. Karssemeijer & te Brake (1998) , and Lau & Bischof (1991) both created a mapping between the left and right breast by using a set of control points defined on the skin line, in a method for asymmetry detection. Karssemeijer and te Brake found a small benefit when using asymmetry as an additional feature in their detection scheme. However, to our knowledge, no studies have been published since then that confirm the usefulness of asymmetry for automated detection of breast lesions. Sajda *et al.* (2002) studied the use of context information by applying hierarchical pyramid neural networks for microcalcification and mass detection. This technique used multi-scale decomposition of an image, via pyramid transforms, and the subsequent integration of multi-scale image features by a hierarchical neural network, to learn contextual relationships from the data. They concluded that this method might be useful for exploiting context for a variety of image analysis problems.

Concluding from the above described studies, apart from the study of Paquerault *et al.*, the improvements of CAD results are either small or not clear. The purpose of this study was to investigate to what extent human readers make use of context information derived from the whole breast area and from asymmetry for the tasks of mass detection and classification. The effect of using two views of the same lesion, like CC and MLO, was not studied. We assume that trained human readers have learned to exploit image context. Results can guide further development of CAD algorithms. To our knowledge, results from human observer studies have not been used before to decide where to focus CAD development.

Some context-related observer studies regarding detection and interpretation in medical images have been published. Kundel and colleagues have found evidence that human observers search radiographs by alternating between a global and focal feature analysis, integrating both sets of features in their decision process (Kundel 2000) . Swensson *et al.* (1982) studied the difference in radiographic interpretation with and without search in chest radiographs. They concluded that there are perceptual mechanisms in the process of visual search that facilitate the observer's separation between normal and abnormal features. Burgess *et al.* (2001) conducted a study where the observers were asked to detect a projected sphere of variable size superimposed on two types of backgrounds: real mammograms and low-pass filtered noise with the same power spectrum. Results suggested that structures of the real mammograms helped the observers in a 2AFC (two-alternative forced-choice) experiment.

In our experiments, we investigated to what extent observers make use of context information by presenting the observers with a number of cases in three different ways, showing only a region of interest, showing the whole image, and showing the whole case. The observers were asked to rate each presented view by giving it a malignancy score. There were two tasks, a detection task, for discriminating between FP and TP regions, and a classification task for discriminating between benign and malignant. Detailed ex-

planation of the experiments can be found in Sec. 7.2.1. Multiple-Reader Multiple-Case ROC analysis was used to evaluate the results (see Sec. 7.2.2). Results are presented in Sec. 7.3.

7.2 Methods

7.2.1 Experimental setup

Use of context information for detection

To study the use of context information on the detection performance, 60 TP and 60 FP CAD detected regions were selected from mammograms from an annotated database containing cases from the Dutch breast cancer screening program. The mammograms were digitized with a Canon CCD scanner at a pixel resolution of 50 μm x 50 μm, and averaged down to a resolution of 100 μm x 100 μm, maintaining the original gray value depth of 12 bits. The false positives were regions detected by our CAD program (Karssemeijer & te Brake 1996; te Brake & Karssemeijer 1999) with a high suspiciousness level. We used a random sampling of regions that were prompted in normal cases, using a threshold level of 0.5 FP/image. The normal cases did not include benign lesions, and were all checked by an experienced radiologist.

The cases were presented in three ways, showing only a region of interest (*roi*), showing the whole image of the breast (*image*), and showing also the same mammographic view of the other breast (*bilat*), see Fig. 7.1. The roi contained the region and a border of 1 cm around the region, and was always positioned in the center of the display regardless of the location of the region in the mammogram. In each trial, a contour indicated the region under investigation. This contour was obtained by segmentation of the region using dynamic programming (Timp & Karssemeijer 2004). Presentation was done on a dedicated mammography workstation (MBC-SCR1, MeVis BreastCare) that was developed by a European Consortium in which we participated [1]. The workstation was equipped with two high-resolution CRT displays (BARCO, MGD 521, 300 Cd/m^2, using BarcoMed 5MP1H 12 bit graphics boards). Every trial started with the available images presented in overview mode (200 μm). The images could subsequently be displayed at full resolution (100 μm). By pressing a button on a dedicated keypad the contour of the region under investigation could be displayed or removed.

The cases and reading modes were randomized, and all trials were divided into three reading sessions of 120 trials. In each session, each of the regions was displayed once in one of the reading modes. On average there were two days between different reading sessions, and each session took roughly an hour. Five observers participated in the

[1] Soft-Copy REading ENvironment SCREEN, EU-Project, IST-1999-10246.

experiment. The observers were all non-radiologists with extensive experience in mammography. The observers were asked to rate the likelihood of malignancy on a continuous scale between 0 and 100. There was no time limit for decision making.

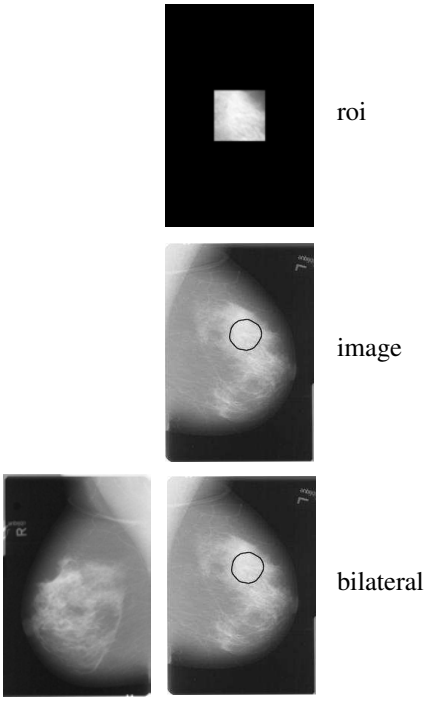


Figure 7.1: *Schematic overview of the reading modes for the detection study.*

Use of context information for classification

The use of context information for classification was studied using 40 malignant and 40 benign regions selected from the above mentioned database. The malignant cases were biopsy proven, and the benign cases were either histologically confirmed or had at least a 6-month follow-up with mammography without suspicion for malignancy. The cases were presented in two ways, showing only a region of interest, if available in both MLO and CC view (*roi*), and showing all available views (*case*), see Fig. 7.2. The contour that indicated the region under investigation was derived from the annotation by an experienced radiologist. Here all trials were divided into 4 sessions of approximately half an hour each. The same observers participated, and other conditions were equal to the detection study.

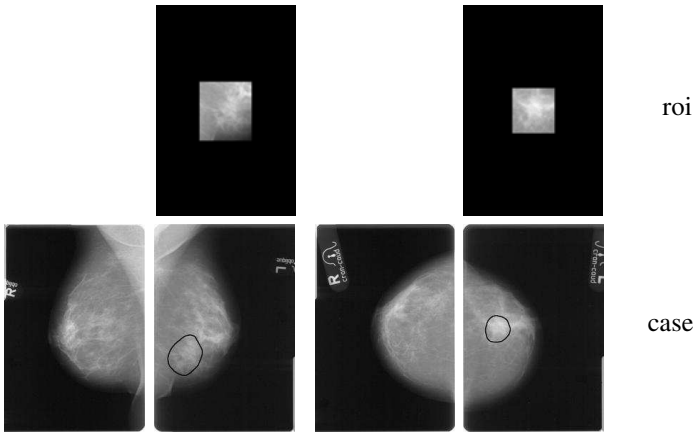


Figure 7.2: *Schematic overview of the reading modes for the classification study.*

7.2.2 Data analysis

The results were analyzed using Receiver Operating Characteristic (ROC) methodology (Metz 1986) and performances were quantified by means of area under the ROC curve (A_z). Statistical analysis was performed using the Dorfman-Berbaum-Metz approach (Dorfman *et al.* 1992). This method has been widely adopted in recent years for analyzing experimental data obtained in a multi-reader multi-case (MRMC) study design. It has the advantage that both reader and case variability are taken into account in a proper way, such that generalization to both the population of readers and cases is permitted. The LABMRMC software[1], which is publicly available, was used for computations. Also CAD stand-alone results were computed for mass detection. The output of our CAD scheme provided a measure of suspiciousness for each region used in the detection experiment. This level was used to generate a ROC curve for the regions used in the experiment.

7.3 Results

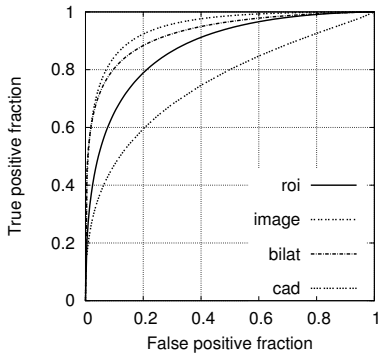
The average ROC curves for the detection task are presented in Fig. 7.3. Also CAD stand-alone results are provided. In Table 7.1 the A_z values for the three reading modes are presented for each observer. All observers showed an improvement in discriminating between TP and FP CAD prompts when using the whole image instead of only the roi

[1] `http://www-radiology.uchicago.edu/krl/roc_soft.htm`

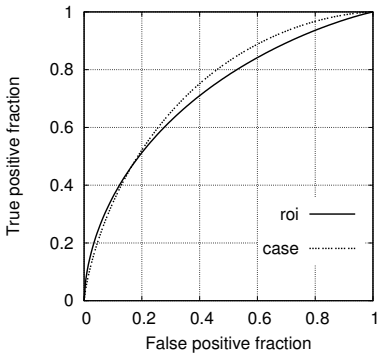
(LABMRMC, significant, 95 % confidence level). Figure 7.3 also shows that CAD standalone performance was significantly worse than that of the observers. Remarkably, the results for the bilateral reading mode were slightly lower than those for the image mode.

The average ROC curves for the classification task are presented in Fig. 7.3. Individual observer A_z values are presented in Table 7.2. Only two observers showed an improvement in discriminating between benign and malignant when using the whole case instead of only the roi. MRMC analysis showed that there was no significant improvement.

A group of 6 radiologists classified the same set of cases, their performance was comparable to the observer performances (average A_z radiologists 0.77, observers 0.74 case based, A_z range radiologists 0.71 - 0.83, observers 0.63 - 0.81). There was no significant difference in performance between the group of radiologists and the observers (Wilcoxon rank sum test, p = 0.6623).



(a) detection task



(b) classification task

Figure 7.3: *Average ROC curves showing the average observer performance in discriminating between FP and TP (a), and benign and malignant (b). The average curves were obtained by averaging the parameters of the individual curves obtained using the LABMRMC software.*

7.3.1 Increase in performance when using context - examples

This paragraph shows some examples of cases where the observers improved their malignancy score when scoring with more context information. In Fig. 7.4 examples are given for the detection task, and in Fig. 7.5 for the classification task.

Table 7.1: *Results of the detection experiment. A_z values and standard error are given.*

Observer	roi	image	bilat
1	0.877 (0.032)	0.947 (0.019)	0.925 (0.027)
2	0.901 (0.030)	0.950 (0.019)	0.931 (0.023)
3	0.937 (0.022)	0.971 (0.014)	0.963 (0.016)
4	0.895 (0.029)	0.949 (0.019)	0.911 (0.029)
5	0.850 (0.036)	0.901 (0.028)	0.877 (0.032)

Table 7.2: *Results of the classification experiment. A_z values and standard error are given.*

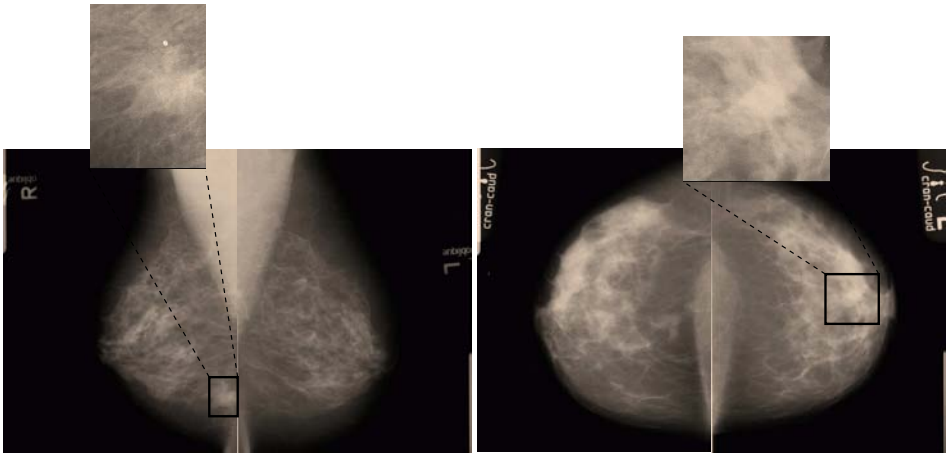
Observer	roi	case
1	0.688 (0.060)	0.724 (0.057)
2	0.689 (0.058)	0.633 (0.062)
3	0.806 (0.049)	0.804 (0.049)
4	0.800 (0.049)	0.811 (0.048)
5	0.576 (0.063)	0.706 (0.057)

7.4 Discussion

Our results show that the observers improved their performance when using context information for classification of false positive and true positive regions. There was no significant difference between their performance in the image and bilateral reading mode.

For classification of benign and malignant lesions the use of context seemed less important. Only two out of five observers showed an improvement in discriminating between benign and malignant when using the whole case instead of only the roi. MRMC analysis showed that there was no significant improvement.

Our observers were non-radiologists with extensive experience in mammography. Since the average performance of a group of radiologists was comparable to the average observer performance for the classification task, we think that it is unlikely that the reading skills of the observers influenced the results. Similarly, Wivell *et al.* (2003) evaluated the ability of radiographers to read screening mammograms. They concluded that radiographers are able to read screening mammograms at least as well as radiologists and do not take longer to do so. However, it can not be ruled out that because of more extensive training radiologists might show a higher performance increase when using context than our group of observers.



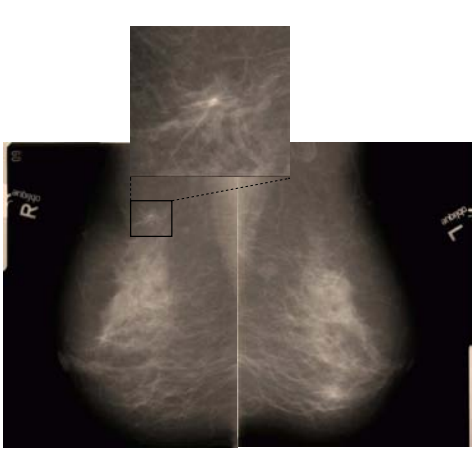
(a) (b)

Figure 7.4: *For the detection task; Two examples where the observers improved their malignancy score when viewing the whole image instead of only a roi containing the region. In (a) a TP region with average observer scores: 31, 54, and 71, for respectively the roi, image and bilateral reading mode. In (b) a FP region with average observer scores: 54, 28, and 31, for respectively the roi, image and bilateral reading mode.*

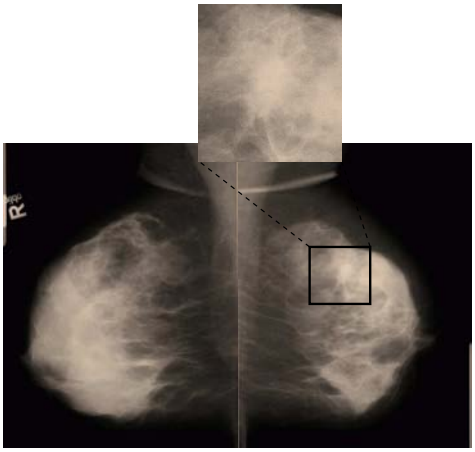
The performance of the observers for the detection task did not improve when using the bilateral reading mode instead of the image reading mode. One might argue that this is because of lack of training. However, since the average performance of a group of radiologists was comparable to the average performance of our observer group, this does not seem to be the most likely explanation. A more probable explanation is that asymmetry is not a very specific sign, and it only plays a role in a very limited number of cases. And as the majority of asymmetric densities are due to normal variation of the parenchymal pattern, this might even explain the small decrease in performance for the bilateral reading mode.

Our results indicate that context information might be used to improve CAD detection performance. However, since our CAD detection performance was still much worse than the average observer performance, it should be concluded that there is still a lot to be gained from improvement of local feature extraction. Note however that this might be different for other CAD systems.

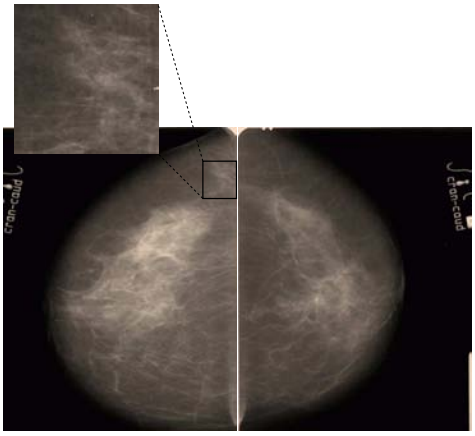
We looked closer at the detection cases where the observers showed a large improvement when using the image reading mode. During a feedback session with the observers



(a) MLO view



(b) MLO view



(a) CC view

Figure 7.5: *For the classification task; Two examples where the observers improved their malignancy score when viewing the whole case instead of only the roi(s) containing the region. In (a) a malignant lesion with average observer scores: 27, and 47, for respectively the roi, and case reading mode. In (b) a benign lesion with average observer scores: 67, and 40, for respectively the roi and case reading mode.*

we tried to find reasons for this improvement. The observers indicated that they used location information and the presence or absence of comparable densities in the breast. They also indicated that they used the relative location of the region with respect to the glandular tissue.

7.5 Conclusions

Context information can be used to improve CAD programs for mass detection. However, there is still a lot to be gained from improvement of local feature extraction and classification. This is demonstrated by the fact that the observers did much better in classifying TP and FP regions than our CAD program.

Experimental results show that the observers were not able to make use of bilateral comparison to improve detection. This might be due to insufficient experience in reading mammograms. It seems more likely however, that asymmetry is only helpful in few cases. For classification of benign and malignant masses context seems to be less important. This suggests that observer experiments in which only regions with abnormalities are presented instead of complete mammograms may provide valid results for this task.

Chapter 8

Volumetric breast density estimation from full field digital mammograms [1]

8.1 Introduction

Breast tissue density has been identified as an important risk factor for breast cancer development. This relationship was first suggested by Wolfe (1976) and later confirmed by others (Byrne *et al.* 1995). Also changes in density may be related to breast cancer risk (van Gils *et al.* 1999a). For details, the reader is referred to a review by Heine (Heine & Malhotra 2002b; Heine & Malhotra 2002a). Tissue that appears as dense on a mammogram is a combination of connective tissue structures and epithelial tissue, together referred to as fibro-glandular. It contrasts with fatty tissue that has a more transparent appearance. Breast tissue density plays an important role in epidemiological studies. If the relation between density and risk is causal, this may have a potential for preventive measures. Breast density assessment may also be used to identify a high risk population, who might benefit from more frequent screening and/or additional imaging with other modalities (van Gils *et al.* 1999b), or in Computer-aided Detection (CAD) systems to increase performance.

Radiologists usually estimate breast density from mammograms in a qualitative way, for example using the density categories of the Breast Imaging Reporting and Data System (BI-RADS) lexicon or some other scheme. There is, however, need for a less subjective and more quantitative measure of breast density that can be obtained automat-

[1] The content of this chapter has been published previously in van Engeland *et al.* (2006).

ically. A number of methods have been published that report automated classification of parenchymal patterns (Caldwell *et al.* 1990; Tahoces *et al.* 1995; Karssemeijer 1998; Wang *et al.* 2003b) or computerized measurement tools for breast density assessment (Byng *et al.* 1994; Byng *et al.* 1998b; Saha *et al.* 2001; Zhou *et al.* 2001). Most of these are based on determination of the size of the dense tissue projection on a mammogram, but some researchers have also explored use of texture measures for risk assessment (Huo *et al.* 2000; Huo *et al.* 2002). If a method for segmentation of dense tissue is used, the (relative) size of the dense tissue projection can be used directly instead of categorizing breast density measurements.

As a quantitative measure of breast density the projected dense tissue area is not ideal, as it is not invariant to compression and projection angle. Although harder to estimate from mammograms, dense tissue volume is a more appropriate measure for breast density than its projected area. Using a physical model of image acquisition, Highnam & Brady (1999) showed how mammograms can be converted to an *interesting tissue* (h_{int}) representation, in which every pixel represents the amount of interesting tissue (fibro-glandular tissue and cancerous tissue). Their normalization method is based on the assumption that the X-ray attenuation coefficients of fibro-glandular and cancerous tissue are nearly equal, but are quite different from that of fatty tissue. After normalization, a mammogram is corrected for scattered radiation and the dependency of image formation parameters like tube voltage, spectrum, and exposure time. A more recent name given to this representation is Standard Mammogram Form (SMF). Recently, Marias *et al.* (2004) presented a study with digitized mammograms that describes application of this SMF representation to measure changes in breast density due to hormone replacement therapy (HRT). Pawluczyk *et al.* (2003) also presented a method for quantitative volumetric analysis of mammographic density from digitized film-screen mammograms. The method is based on initial calibration of the imaging system and correction for variations in exposure factors and film processing characteristics. It requires a small calibration wedge to be imaged in every mammogram and imaging conditions and compressed breast thickness should be known. Results obtained with phantoms showed that density measurements could be made within 5 % from the actual value. No clinical validation was performed yet.

Most breast density estimation studies to date focus on the use of digitized mammograms. When using films, it remains cumbersome to obtain reliable calibration data. In particular, sensitometric variables may change slightly from day to day.

Therefore, even when films are digitized it remains hard to design a reliable measurement procedure for breast density in clinical practice. The introduction of digital mammography, however, gives rise to new opportunities. Stability and linearity of digital detectors, and the fact that many parameters for calibration are stored with the image itself, will make breast density estimation much more feasible. In this study we investigate a fully automated physics-based method to determine the volume of glandular tissue from full field digital (FFD) mammograms. The method takes into account different energy spectra for different anode target/filter materials and is designed to work on raw (unprocessed) data.

To validate breast density measurement methods it has been common practice to use results obtained by visual assessment of radiologists as a gold standard. Obviously, this approach suffers from the inaccuracies of visual judgments that one tries to avoid by using automated image analysis. In some more recent studies, breast MRI data has been used (Stoutjesdijk & Karssemeijer 2000; Wang *et al.* 2003a; Wei *et al.* 2004), which allows more accurate and more objective validation. Also in this paper we will use this method. An interactive method for segmentation of fibro-glandular tissue in 3D MRI data sets was designed which was used to determine volume of dense tissue. Dense tissue volumes obtained from MRI data were used as ground truth against which mammogram measurements were compared. Validation was performed using FFD cases acquired with a GE Senographe 2000D. We compared volume estimates in 88 mammographic views from 22 patients.

8.2 Volumetric breast density estimation

In this section we show how the volume of dense tissue can be computed from an unprocessed digital mammogram and how total breast volume can be obtained using a geometric model. For breast density estimation, a polychromatic image model is used that incorporates the photon energy spectrum for different anode target/filter materials and the energy dependency of linear attenuation coefficients of breast tissue. While the model is basically polychromatic, the relation we derive can be written in a monochromatic form with effective linear attenuation coefficients that depend on acquisition settings. All required parameters are obtained from DICOM headers of the digital mammograms and from empirical data in literature. To facilitate implementation of the density estimation procedure, we present parameters for a range of acquisition settings in a table.

8.2.1 Imaging model

The X-ray exposure $I(\mathbf{r})$ at a detector location $\mathbf{r}$ is modeled by a polychromatic model

$$\frac{I(\mathbf{r})}{I_0} = \int_{E=0}^{\infty} p(E)\ e^{-\int_{z=0}^{h(\mathbf{r})} \mu(\mathbf{r},z;E)\,dz} dE, \tag{8.1}$$

with $p(E)$ the normalized photon energy spectrum, which we assume to be spatially homogeneous. In this paper, the spectrum will be fully characterized by anode material, filter material, filter thickness, and the tube voltage. We assume that the detector is calibrated for inhomogeneities due to, e.g., the Heel effect. $\mu(\mathbf{r},z;E)$ is the linear attenuation coefficient for tissue at the three-dimensional location $(\mathbf{r},z)$. It depends on the kind of tissue at location $(\mathbf{r},z)$, but also on the photon energy E. The integral over $\mu(\mathbf{r},z;E)$ is along the line segment with length $h(\mathbf{r})$ of the X-ray beam intersecting the breast tissue.

The normalized photon energy spectrum, $p(E)$, is obtained from (Boone *et al.* 1997), for molybdenum, rhodium, and tungsten anode target materials. The effect of filtration has been modeled using attenuation data provided in (Hubbell & Seltzer 1995). It is assumed that the required acquisition parameters are known. With full field digital mammography these are generally made available as DICOM header tags.

With regard to X-ray attenuation, two types of tissue can be distinguished in the breast, dense and fatty tissue, with linear attenuation coefficients $\mu_d(E)$ and $\mu_f(E)$, respectively (throughout the paper subscripts d and f will be used for dense and fatty tissue). It has been found that in good approximation mammographic imaging can be modeled with only these two tissue components, apart from microcalcifications (Highnam & Brady 1999). For the purpose of breast density estimation, it is safe to ignore microcalcifications, because these, if present, only occupy a very small fraction of the breast. We use tabulated experimental data for linear attenuation coefficients of dense and fatty tissue presented by Byng *et al.* (1998a). The data labeled by PCJ (from (Johns & Yaffe 1987)) are used throughout the paper. To obtain attenuation coefficients at a given energy the experimental data were interpolated. We use cubic splines to interpolate $\ln \mu_f(E)$ and $\ln \mu_d(E)$. This function appeared to be almost linear in the parameter range of interest.

To simplify computations, we compute effective attenuation coefficients for fatty and dense tissue as a function of the anode and filter material, tube voltage, and tissue thickness h. The attenuation of a mixture of dense and fatty tissue with thicknesses h_d and h_f, at a given location, is given by

$$\frac{I}{I_0} = \int_{E=0}^{\infty} p(E)\ e^{-\mu_f(E)h_f - \mu_d(E)h_d} dE. \tag{8.2}$$

It appears that for typical spectra used in mammographic imaging this attenuation can very well be approximated by an exponential function. This implies that we can write

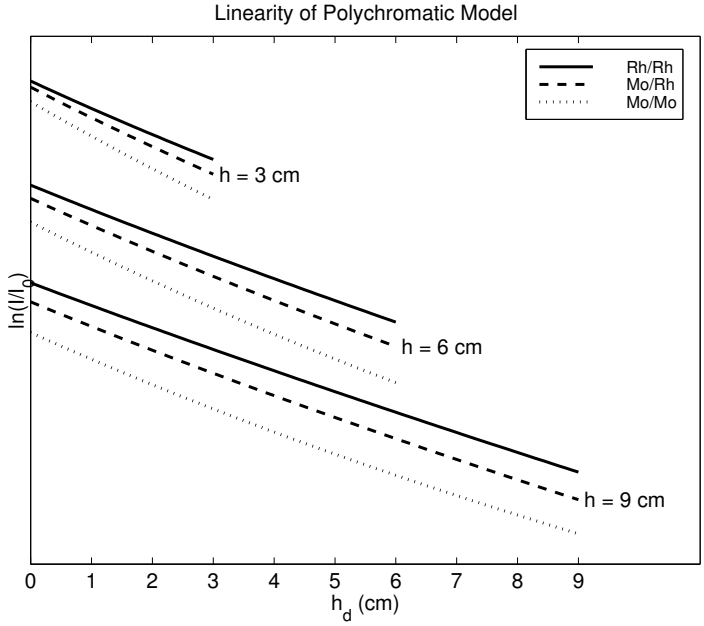


Figure 8.1: *For typical spectra used in mammography the logarithm of the attenuation is approximately linear with dense tissue thickness, for a given breast thickness. This is illustrated in the figure for an imaging system with different anode/filter combinations at a tube voltage of 27 kVp.*

the logarithm of the attenuation as

$$\begin{aligned} \ln \frac{I}{I_0} &\approx -\mu_{f,eff}\, h_f - \mu_{d,eff}\, h_d \\ &= -\mu_{f,\,eff}\, (h - h_d) - \mu_{d,eff}\, h_d, \end{aligned} \tag{8.3}$$

where the effective attenuation coefficients depend on acquisition parameters, and with h the breast thickness. The linearity is illustrated in Fig. 8.1 for a tube voltage of 27 kVp and different anode/filter combinations. It is noted that the dependence on breast thickness includes the beam-hardening effect.

Effective linear attenuation coefficients for dense and fatty tissue are estimated by linear regression of Eq. 8.3 for a range of acquisition parameters. In the next section it will be shown that for computation of dense tissue volume only the difference of effective linear attenuation coefficients, $\mu_{d,eff} - \mu_{f,eff}$, is needed. In Table 8.1 these differences are given for a range of acquisition parameters.

Table 8.1: *Differences of effective linear attenuation coefficients, $\mu_{d,eff} - \mu_{f,eff}$, under several acquisition conditions. For different anode target materials, filter materials, breast thicknesses h, and $kVps$, the effective linear attenuation coefficients are computed with the polychromatic model (Eq. 8.3). The filter thickness was $30\,\mu m$ in all cases.*

Mo/Mo	$h\ (cm)$							
kVp	2	3	4	5	6	7	8	9
24	0.513	0.477	0.452	0.433	0.417	0.403	0.390	0.379
25	0.497	0.461	0.435	0.414	0.396	0.379	0.363	0.348
26	0.484	0.448	0.421	0.398	0.378	0.358	0.340	0.324
27	0.468	0.432	0.403	0.377	0.354	0.333	0.313	0.295
28	0.456	0.419	0.388	0.360	0.335	0.312	0.291	0.273
29	0.442	0.403	0.370	0.340	0.313	0.289	0.267	0.249
30	0.429	0.387	0.352	0.320	0.292	0.267	0.246	0.228
31	0.414	0.371	0.334	0.301	0.272	0.247	0.227	0.210
32	0.402	0.357	0.319	0.285	0.256	0.232	0.212	0.196
Mo/Rh								
24	0.448	0.415	0.392	0.373	0.358	0.346	0.335	0.326
25	0.432	0.401	0.379	0.361	0.345	0.333	0.322	0.313
26	0.422	0.392	0.370	0.352	0.336	0.323	0.312	0.303
27	0.413	0.384	0.361	0.342	0.327	0.313	0.301	0.291
28	0.407	0.377	0.354	0.335	0.318	0.303	0.291	0.280
29	0.399	0.369	0.345	0.325	0.307	0.291	0.277	0.265
30	0.392	0.361	0.336	0.314	0.295	0.278	0.263	0.249
31	0.384	0.352	0.326	0.302	0.282	0.263	0.247	0.233
32	0.377	0.344	0.316	0.291	0.269	0.250	0.233	0.219
Rh/Rh								
24	0.429	0.393	0.369	0.351	0.337	0.326	0.318	0.310
25	0.406	0.373	0.351	0.335	0.322	0.313	0.304	0.298
26	0.388	0.358	0.338	0.323	0.312	0.302	0.295	0.288
27	0.374	0.346	0.327	0.313	0.302	0.293	0.285	0.278
28	0.363	0.337	0.319	0.305	0.294	0.285	0.277	0.269
29	0.353	0.328	0.310	0.297	0.286	0.276	0.267	0.259
30	0.344	0.320	0.303	0.289	0.277	0.267	0.258	0.249
31	0.336	0.312	0.295	0.281	0.268	0.257	0.247	0.238
32	0.328	0.304	0.286	0.272	0.259	0.247	0.236	0.226

8.2.2 Dense tissue volume

In an unprocessed FFD mammogram pixel values are proportional to the total exposure $I(\mathbf{r})$. Hence, the image model is obtained from Eq. 8.1 by replacing exposure value (I) by pixel value (g)

$$\begin{aligned}\frac{g(\mathbf{r})}{g_0} &= \int_{E=0}^{\infty} p(E)\; e^{-\mu_f(E)\,h_f(\mathbf{r})-\mu_d(E)\,h_d(\mathbf{r})}dE \\ &= \int_{E=0}^{\infty} p(E)\; e^{-\mu_f(E)\,h(\mathbf{r})-(\mu_d(E)-\mu_f(E))\,h_d(\mathbf{r})}dE. \end{aligned} \tag{8.4}$$

In this equation the photon energy spectrum $p(E)$, and the attenuation coefficient $\mu_f(E)$ and $\mu_d(E)$ are known from empirical data. When we could also determine breast thickness $h(\mathbf{r})$ and the pixel value associated with the incident X-ray beam g_0, computation of the dense tissue thickness $h_d(r)$ would be straightforward. Unfortunately, it is not easy to obtain accurate estimates of these parameters in practice. A compressed breast thickness measurement is provided by most modern mammographic imaging systems, but measured values are usually not very precise and small errors in breast thickness lead to large deviations in $h_d(r)$ when computed directly with Eq. 8.4. Moreover, breast thickness is not uniform and falls off in the periphery of the breast. This should be taken into account. Also the background exposure may be hard to obtain. Due to its high value, the background in mammograms is often clipped, in which case the background pixel value does not give a representative value for I_0. Because of these problems, we use another method to determine $h_d(r)$ which will be explained below.

Suppose that we can carry out a thickness correction transform on the mammogram in which a layer of adipose tissue with attenuation coefficients $\mu_f(E)$ and thickness $H - h(\mathbf{r})$ was added to the breast. Methods to perform such a transform have been described in the literature, e.g. (Bick *et al.* 1996; Byng *et al.* 1997; Snoeren & Karssemeijer 2004; Snoeren & Karssemeijer 2005), and application of these has become more or less standard to enhance visualization of the periphery of the breast. Ideally, after thickness correction we would obtain an image in which the following equation would hold

$$\frac{\overline{g}(\mathbf{r})}{g_0} = \int_{E=0}^{\infty} p(E)\; e^{-\mu_f(E)\,H-(\mu_d(E)-\mu_f(E))\,h_d(\mathbf{r})}dE. \tag{8.5}$$

In this image, pixel values only vary with dense tissue thickness. Purely adipose tissue looks the same everywhere in the corrected mammogram, no matter the local breast thickness. In most mammograms it is easy to identify a location that corresponds with almost pure fatty tissue. Because the attenuation coefficient of fatty tissue is smaller than that of dense tissue, this location is determined by the maximum of $\overline{g}(\mathbf{r})$.

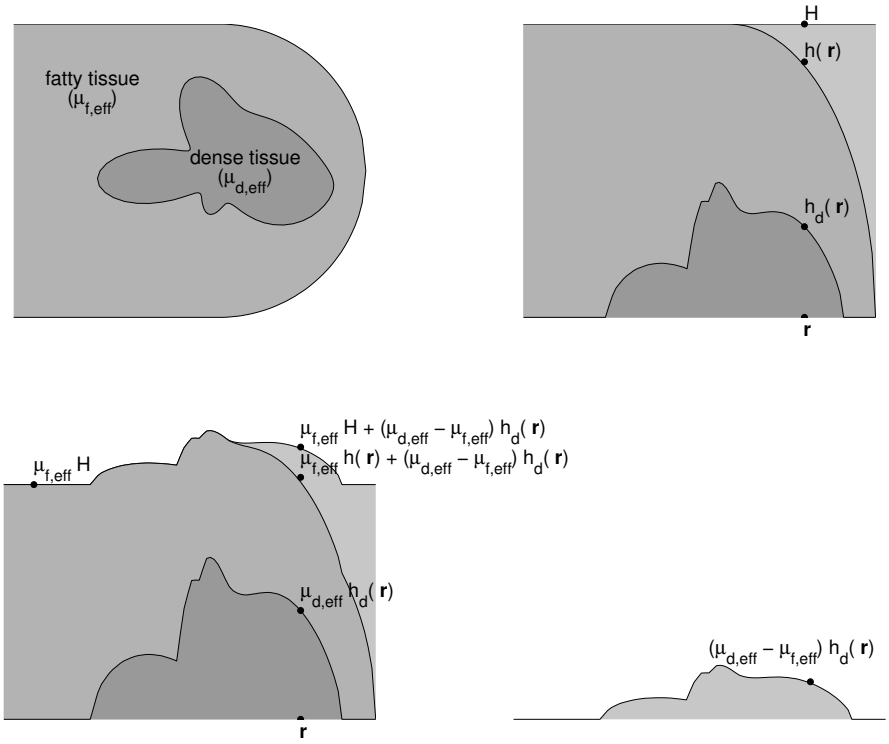


Figure 8.2: *Sketch of volumetric breast density estimation. The upper left panel shows a cross section of a breast with one lump of dense tissue. The upper right panel shows the thickness profiles for this cross section; from bottom to top: thickness of dense tissue, $h_d(\mathbf{r})$; breast thickness, $h(\mathbf{r})$; and breast thickness after thickness correction, H. The lower left panel shows the corresponding exposure profiles for different attenuation coefficients after linearization. The lower right panel shows the bump in the lower left panel, from this the thickness profile of dense tissue is obtained.*

The corresponding image model for purely fatty tissue is obtained by setting $h_d(\mathbf{r}) = 0$ in Eq. 8.5:

$$\frac{\overline{g}_f}{g_0} = \int_{E=0}^{\infty} p(E)\ e^{-\mu_f(E)\,H} dE, \tag{8.6}$$

where we neglected the skin (we will elaborate further on its effect in the discussion). By substituting the pixel value of fatty tissue $\overline{g}_f$ in 8.5 we obtain

$$\frac{\overline{g}(\mathbf{r})}{\overline{g}_f} = \frac{\int_{E=0}^{\infty} p(E)\ e^{-\mu_f(E)\,H-(\mu_d(E)-\mu_f(E))\,h_d(\mathbf{r})} dE}{\int_{E=0}^{\infty} p(E)\ e^{-\mu_f(E)\,H} dE}. \tag{8.7}$$

In principle, $h_d(\mathbf{r})$ can be solved from this equation if H is known. It turns out, however, that the value of H is not critical anymore, due to the internal calibration with a fatty tissue pixel value. In fact, when we apply the exponential approximation of Sec. 8.2.1 and rewrite 8.7 with the effective attenuation coefficients $\mu_{f,eff}$ and $\mu_{d,eff}$ the explicit dependency of H disappears, and we obtain the following relation

$$\begin{aligned} \frac{\overline{g}(\mathbf{r})}{\overline{g}_f} &= e^{-(\mu_{d,eff}-\mu_{f,eff})\,h_d(\mathbf{r})} \Rightarrow \\ h_d(\mathbf{r}) &= -\frac{1}{\mu_{d,eff}-\mu_{f,eff}} \ln \frac{\overline{g}(\mathbf{r})}{\overline{g}_f}. \end{aligned} \tag{8.8}$$

It should be noted, however, that the dense tissue thickness remains dependent on breast thickness as the effective attenuation coefficients vary with H. The total volume of glandular tissue is given by

$$V_d = \int_{\mathrm{B}} h_d(\mathbf{r})\ d^2\mathbf{r} = -\frac{1}{\mu_{d,eff} - \mu_{f,eff}} \int_{\mathrm{B}} \ln \frac{\overline{g}(\mathbf{r})}{\overline{g}_f}\, d^2\mathbf{r}, \tag{8.9}$$

where the integral is taken over the projected breast area B, excluding the pectoral area. Figure 8.2 illustrates how the dense tissue thickness is obtained.

In the implementation we use a thickness correction method developed previously (Snoeren & Karssemeijer 2005) to obtain $\overline{g}$. This is an iterative method, in which anisotropic diffusion of the fatty tissue region is used to obtain a smooth surface representing the shape of the breast. Dense tissue is interpolated along curves running in parallel to the breast edge. The breast and the pectoral region are segmented using a method described previously (Karssemeijer 1998). To determine the fatty tissue reference value g_f we determine the maximum pixel value in the interior of the breast using a large quantile (0.99) of the pixel value histogram. Only for extremely dense breasts, we compute this quantile on the whole peripherally enhanced breast, because in those cases the interior may not contain pixels in which only fatty tissue is mapped.

8.2.3 Total breast volume

To compute relative breast density the total volume of the breast needs to be computed. Using the exponential approximation (similar to approximation from Eq. 8.2 to Eq. 8.3) and effective attenuation coefficients (see also Fig. 8.2), Eqs. 8.4 and 8.5 can be approximated as

$$\frac{g(\mathbf{r})}{g_0} = e^{-\mu_{f,eff}\, h(\mathbf{r}) - (\mu_{d,eff} - \mu_{f,eff})\, h_d(\mathbf{r})}, \text{ and} \tag{8.10}$$

$$\frac{\overline{g}(\mathbf{r})}{g_0} = e^{-\mu_{f,eff}\, H \;\; -(\mu_{d,eff} - \mu_{f,eff})\, h_d(\mathbf{r})}. \tag{8.11}$$

Using these equations the thickness of the breast $h(r)$ follows from $\ln \frac{\overline{g}(\mathbf{r})}{g(\mathbf{r})}$, and the total breast volume can be obtained as follows

$$V = \int_{\mathrm{B}} h(\mathbf{r})\; d^2\mathbf{r} = \int_{\mathrm{B}} (H + \frac{1}{\mu_{f,eff}} \ln \frac{\overline{g}(\mathbf{r})}{g(\mathbf{r})})\, d^2\mathbf{r}. \tag{8.12}$$

However, in this study we use a geometric approach which has the advantage that it does not depend on the effective attenuation coefficient of fatty tissue. In the geometric approach, perpendicular cross sections of the peripheral zone of the breast are approximated by semi-circles, and the interior by parallel planes. Then breast thickness as a function of location is given by

$$h(\mathbf{r}) = \begin{cases} 2\sqrt{\left((H/2)^2 - (H/2 - d(\mathbf{r}))^2\right)}, & d(\mathbf{r}) < H/2; \\ H, & d(\mathbf{r}) \geq H/2, \end{cases} \tag{8.13}$$

where H is the compressed breast thickness and where $d(\mathbf{r})$ is the Euclidean distance to the edge between breast tissue and background in the mammogram. To compute this we perform a distance transform of the segmented breast area (Borgefors 1986). In order to avoid inaccuracies in larger breasts, where the skinline sometimes intersects the image border, we extrapolate the segmented tissue region outside the image matrix before applying the distance transform. This is partly done by straight lines and partly by segments of ellipses, depending on the location of the extrapolation and on the view.

8.3 Validation

To validate our method for volumetric density estimation from mammograms we collected a series of cases in which both digital mammography and breast MRI had been performed. For this series, volumes of dense breast tissue were determined in the 3D

breast MRI exams and compared with the results of the mammographic density estimation procedure. To determine volume of glandular tissue in MRI we used a semi-automatic segmentation, which is outlined in Sec. 8.3.2. On a larger set of cases, we also compared the amounts of glandular tissue determined from craniocaudal (CC) and medio lateral oblique (MLO) projections of the same breast, which should be the same.

8.3.1 Data set

In our institute, breast MRI is used frequently for screening in a group of younger women with high familial or genetic risk (Kriege *et al.* 2001). From this group of patients we selected 22 cases for which both breast MRI data and a full field digital mammogram were available. Time between the two exams was less than one year for all cases, and two months on average. All exams had been performed between December 2000 and September 2003.

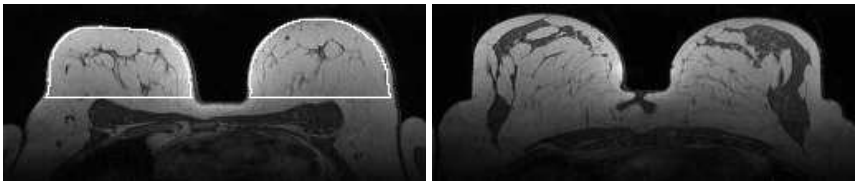
The digital mammograms used in the study were acquired on a GE Senographe 2000D using standard clinical settings, including the use of an anti-scatter grid. Breast MRI examinations were performed on a 1.5 T system (SIEMENS 1.5 T, MAGNETOM VISION), with a dedicated breast coil (CP Breast Array, Siemens, Erlangen). A dynamic contrast enhanced T1-weighted FLASH-3D sequence was used, with repetition time of 8.1 ms, an echo time of 4 ms, and a flip angle of 20 degrees. The pixel spacing was 1.25 mm x 1.25 mm, and the slice thickness 1.5 mm. Per series 108 slices were acquired, without interslice gap. Patients were scanned in prone position. The pre-contrast series was used for the segmentation of glandular tissue.

To compare the glandular tissue volumes determined from the CC and the MLO view of the same breast we could use a much larger data set, since no corresponding MRI data was required. For this comparison we used a series of 1820 mammograms, also acquired on a GE Senographe 2000D, containing 910 CC/MLO pairs.

8.3.2 Segmentation of dense tissue in MRI

To segment glandular tissue in the MRI data we used a semi-automatic approach. Seen in the axial view, we used a plane perpendicular to the breasts to separate the breasts from the chestwall. In Fig. 8.3(a) an example is given of the selected breast volume seen in an axial slice of the MRI examination. This approach is bound to underestimate the total breast volume. However, selecting voxels closer to the chest increases the risk of making large errors, because part of the pectoral muscle might then be segmented along with the glandular tissue. As our goal was to estimate the glandular tissue volume, underestimation of the total breast volume in the MRI volume was not a disadvantage. In practice, when estimating relative breast density the total breast volume from the mammogram will be used. Therefore, we chose this relatively simple approach, and did not select

cases where the glandular tissue distribution did not allow separation from the chest using a single plane. An example of a case that was not included is shown in Fig. 8.3(b). Here it is obvious that our MRI segmentation method would miss a large part of the glandular tissue. The following paragraphs describe the method to segment the glandular tissue in the MRI volume in detail.



(a) (b)

Figure 8.3: *Selected breast volume indicated on an axial slice of the MRI volume (a). Axial slice of an MRI examination that was not included in the data set because of the distribution of the glandular tissue (b).*

The MRI examinations consisted of 108 coronal slices. To find the first slice containing breast tissue (near the nipple), for every slice the average gray value was determined in a horizontal band of 2 cm (see Fig. 8.4). This first slice S_0 was selected by searching for the maximum of the second derivative of the average gray value. The last slice of interest S_N, at the chestwall, was determined in a similar way by thresholding the average gray value in a vertical band in between both breasts.

For every coronal slice between S_0 and S_N the outline of each breast was determined separately using a dynamic programming segmentation algorithm (Timp & Karssemeijer 2004). This algorithm requires a seed point near the center of the region to be segmented. Seed points were determined by taking the maximum pixel value in the central zone of each breast, after heavily smoothing the image.

Because of the inhomogeneity of the breast coil sensitivity, the signal intensity in the breast region was not uniform across the field of view. We corrected for this by using a first order bias-field correction. We assumed a linear bias-field in the axial direction. The parameters for this function were estimated by minimizing the peak widths of both the fatty and glandular tissue peaks (Full-Width-at-Half-Maximum) by using Powell optimization (Press *et al.* 1992).

For determination of the volume of glandular tissue from the MRI data we used an interactive image display environment in which the user could visualize the 3D MRI data sets with the segmented glandular tissue projected as a color overlay. Thresholding

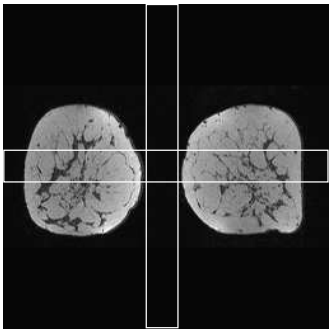


Figure 8.4: *A coronal slice of a breast MRI volume. Horizontal and vertical bands are used to automatically determine the first slice (near the nipple) and the last slice (near the chest) of a volume of interest for glandular tissue segmentation.*

was used to segment glandular from fatty tissue. The users could interactively adapt the threshold value and check the selected pixels by going through the coronal slices. For each case, an optimal threshold value was determined by two observers independently, without having access to the mammograms or the outcome of mammographic density estimation. For all slices between S_0 and S_N the pixels inside the breast area were classified as glandular tissue when the gray value of these pixels was lower than the determined threshold. The glandular tissue volume was estimated by using pixel spacing and slice distance obtained from the DICOM headers of the data sets. For the final result, the volumes of glandular tissue obtained using the two thresholds determined by the observers were averaged.

8.3.3 Comparison of MRI and FFD mammogram estimates

To validate our volumetric breast density estimation from digital mammograms, a number of comparisons with MRI results were made. First, for every available mammographic view (n = 88) the estimated dense tissue volume was compared with the corresponding MRI volume. Second, for each patient (n = 22) average glandular tissue volumes were calculated and compared. Third, to compare the fraction of dense tissue per patient, the dense tissue volume of each patient was divided by the total breast volume. As our approach to segment glandular tissue from the MRI data is bound to give an underestimation of the total breast volume, for both ratios the breast volume determined from the mammogram was used.

We investigated to what extent use of a model to adaptively compute attenuation coefficients based on the acquisition settings and breast thickness contributed to the accuracy

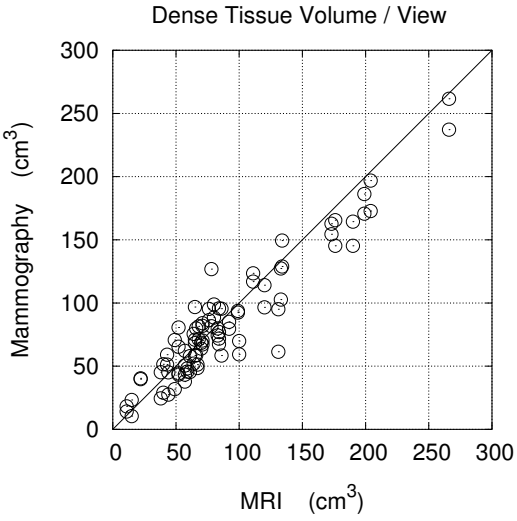
of the results. For that purpose, we also computed results using a simple monochromatic model with fixed values of the attenuation coefficients for $\mu_{f,eff}$ and $\mu_{d,eff}$. These fixed values were obtained by averaging the coefficients of the individual cases in our data set. Comparison with MRI results is presented using the per patient analysis. We also investigated if it is possible to apply our method to processed mammograms, using the *for presentation* images prepared by the GE Senographe. It appeared that the processing applied by GE allows application of our method, because in the interior of the breast a fixed conversion is used by the manufacturer to transform raw pixel values to pixel values in the processed image. We used the transformation experimentally determined by Burgess (2004) for this purpose.

8.4 Results

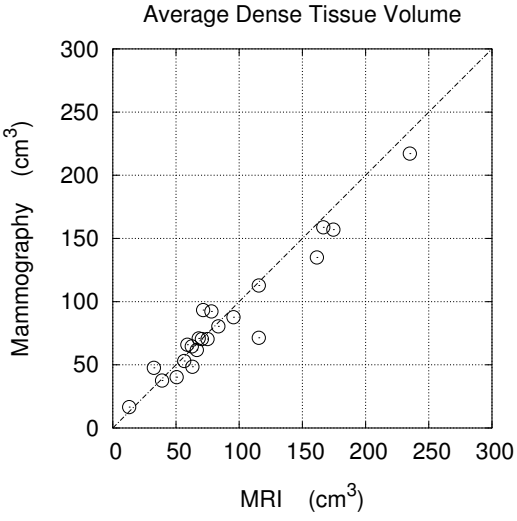
The results of the comparison between dense tissue volumes determined from mammograms and from MRI data are presented in Fig. 8.5. In Fig. 8.5(a) the results are presented of the comparison of the glandular tissue volume determined for every available mammographic view with the corresponding MRI volume. Figure 8.5(b) presents the results per patient. Here the glandular tissue volume per patient was calculated by averaging over the CC and MLO view and over the left and right breast. Figure 8.6 presents the glandular tissue ratios as determined from the mammograms and from the MRI data, also per patient. Both ratios were calculated by dividing by the total breast volume as determined from the FFD mammograms. We computed the relative error in the estimated dense tissue volume for each patient by dividing the absolute value of the difference between volumes obtained with MRI and mammography by the dense tissue volume determined with MRI. The average relative error was 13.6 %.

The results show that there is a high correlation between the dense tissue volumes determined from the mammograms and the MRI volumes. For the comparison of the volumetric density determined for every available mammographic view and the corresponding volumetric density from MRI (see Fig. 8.5(a)) the correlation coefficient is 0.94. When determined per patient, the correlation between the volume of glandular tissue from the mammograms and from the MRI data is 0.97 (see Fig. 8.5(b)). The graph in Fig. 8.7(a) shows the results per patient determined with fixed attenuation coefficients. Figure 8.7(b) shows the results obtained by using processed mammograms. When comparing density volumes with the MRI volumes the method performed significantly better using the raw images (paired t-test one-tailed on differences between volumetric amount of glandular tissue from MRI and FFD mammogram, P = 0.02). The average relative error was 27.8 % and 17.8 %, respectively, when using fixed attenuation coefficients and processed images.

To compare the glandular tissue volumes determined on the digital mammograms for



(a)



(b)

Figure 8.5: *Comparison of dense tissue volumes from MRI and FFD mammograms for individual views, both CC and MLO, (a) and per patient (b).*

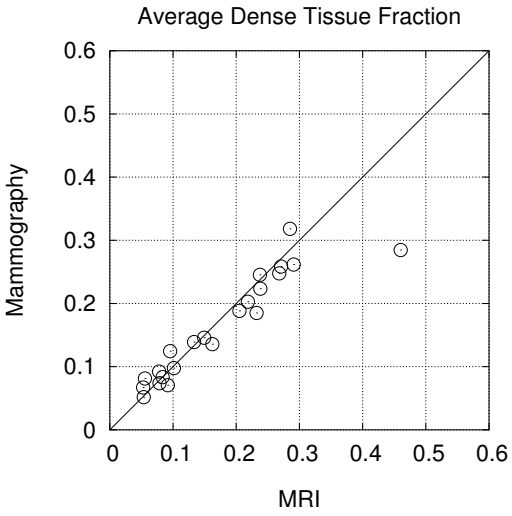


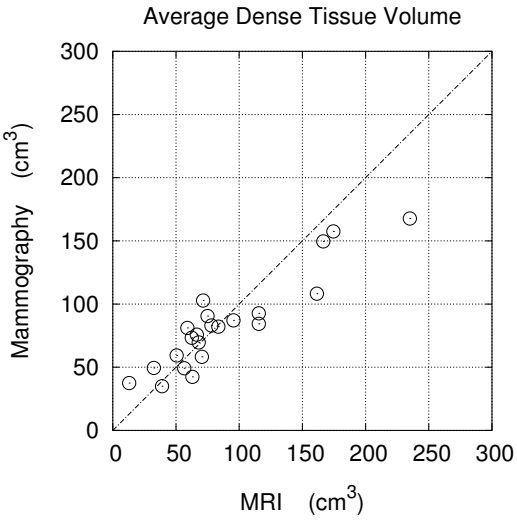
Figure 8.6: *Comparison of average dense tissue fractions per patient, taken relative to breast volumes determined from the mammograms.*

CC and MLO views a much larger data set could be used. Figure 8.8 shows the glandular tissue volumes determined from CC and MLO views of the same breast for 910 CC/MLO pairs. Figure 8.8(a) shows a high correlation between the volumes determined from CC and MLO views (correlation coefficient 0.92). The volume of dense tissue determined from MLO views appears to be slightly lower than the volume determined from CC views (paired t-test one-tailed, $p << 0.001$, see also Fig. 8.8(b)).

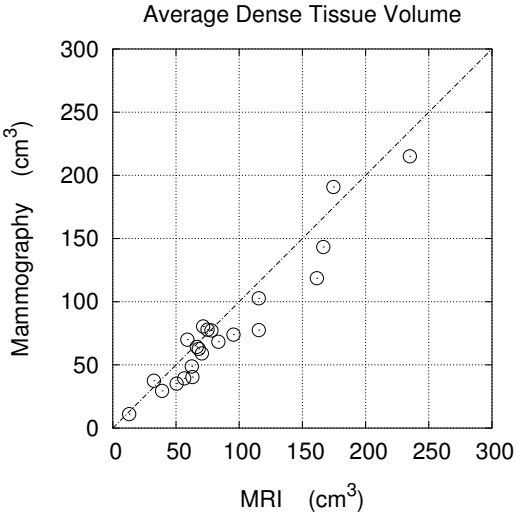
Some examples of mammograms with corresponding breast density volumes are shown in Fig. 8.9. All images are MLO views of the right breast. The mammogram on the right was from the most dense case in our study sample, which was the only case in which our method failed to estimate breast density volume accurately.

8.5 Discussion & conclusions

Results in Fig. 8.5 show that our method provides reliable estimates of the volume of dense tissue in the breast from unprocessed digital mammograms. In particular when evaluation is performed on a per patient basis, where the average breast density volume over four mammographic views is taken, correlation with the MRI ground truth values is high. In practice, interest in breast density is mostly related to breast cancer risk assessment. For that purpose, per patient estimates are more relevant than estimates from

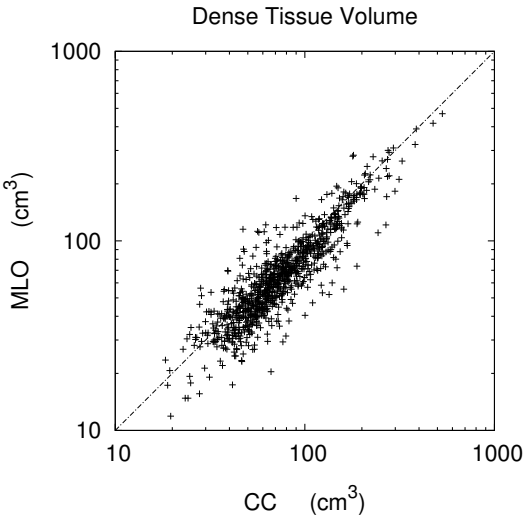


(a)

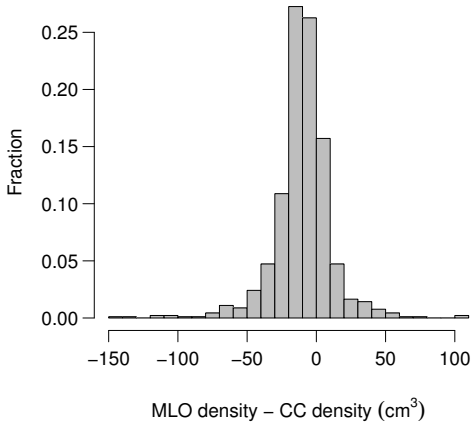


(b)

Figure 8.7: *Average dense tissue volumes per patient obtained by using fixed values of the linear attenuation coefficients for fatty and glandular tissue (a), and by using processed instead of raw digital mammograms (b).*



(a)



(b)

Figure 8.8: *Comparing glandular tissue volume from CC and MLO views for the large data set with raw images (a). Histogram of the difference between volumetric density from MLO and from CC views (b).*

separate mammographic views. The average relative error of the per patient measurements we obtained was 13.6 %, using dense tissue volumes estimated from MRI data as the gold standard.

Our method is based on a polychromatic model. Per image the following parameters are required: Tube voltage, anode and filter materials, and breast thickness. Depending on these parameters, effective attenuation coefficients for fatty and dense tissue are computed, which are subsequently used in a monochromatic computation of dense tissue volume. It was investigated if this approach was more accurate than a simple non-adaptive estimation method, in which fixed values of the attenuation coefficients were used. Results were found to be significantly worse (Fig. 8.7(a)), with an average relative error increasing from 13.6 % to 27.8 %, demonstrating the need for an adaptive approach. This is not surprising, as the variation of the attenuation difference of fatty and dense tissue with the exposure settings and with breast thickness is rather large, as can be observed in Table 8.1. It is noted that filtration thickness was not included as a parameter in our method, due to the fact that this thickness is fixed in the mammography system we used.

To apply the breast density estimation method presented in this paper pixel values must be proportional to exposure. This requirement may exclude the possibility to apply the method to images processed by the manufacturer. Especially if the processing involves some kind of adaptive contrast enhancement, breast density estimation will be severely hampered or rendered impossible. This is not the case with mammograms processed by the GE Senographe 2000D used in this study. The GE processing consists of a peripheral enhancement algorithm and a fixed pixel-to-pixel transform. The latter can be inverted to obtain the original raw pixel values in regions that were not touched by the peripheral enhancement. We used this inverted transform to apply our density estimation method to mammograms processed for presentation, motivated by the fact that mammograms are usually archived in this form. It was found that results were significantly worse than those obtained from the raw images. This is due to the peripheral enhancement algorithm applied by GE. Two effects were found. In the first place it was observed that the GE processed images still had a small breast thickness decrease in the periphery of the breast. This can lead to some underestimation of breast density volume, which can be observed in Fig. 8.7. However, for some cases it may also increase estimated breast density, when the fatty tissue location used for calibration lies in the breast periphery. A second effect was found in some very dense breasts, where the peripheral enhancement algorithm of GE appeared to affect fatty tissue regions in the interior of the breast, which also leads to inaccurate calibration of our method.

For a series of 910 pairs of CC and MLO views we compared fibro-glandular tissue volumes determined in both views (Fig. 8.8). It was found that the measurements correlated well, but that volume estimates were somewhat higher on average in the CC views. Possible causes were investigated. It turned out that it frequently occurred that

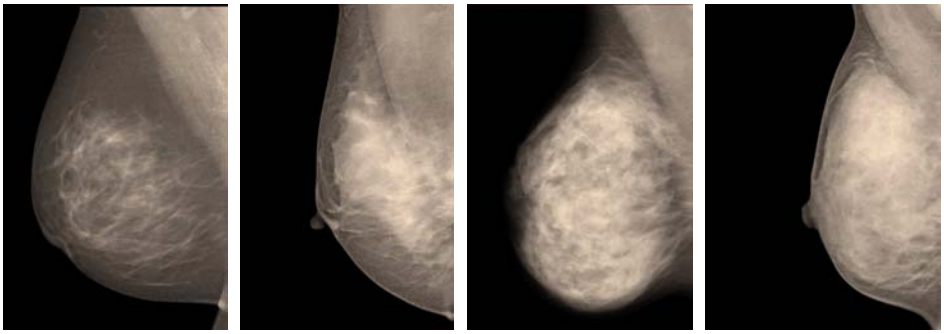


Figure 8.9: *Some examples with increasing breast density from left to right. All images are MLO views of the right breast. Dense tissue volume fractions of these cases determined from MRI are, respectively, 0.08, 0.20, 0.28, and 0.45.*

a small part of the pectoral muscle was projected in the CC views. Our segmentation method only segments the pectoral muscle in MLO views. Therefore, we were not able to exclude the pectoral muscle area when determining the density in CC views. This may explain the small deviation we found.

In our data set we had one outlier, which turned out to be a very dense mammogram. In Fig. 8.6 it is shown that the volume fraction of dense tissue in this case was approximately 45 %. An image of the right MLO view of this case is displayed in Fig. 8.9. Because it is very hard to obtain a reliable calibration location with only fatty tissue, the dense tissue content is underestimated. Our method has to rely on pixel values near the projected skinline for calibration, and these are susceptible for inaccuracy of the peripheral enhancement procedure, which itself is also more difficult to carry out in very dense mammograms. Fortunately, only few mammograms are extremely dense. In another example of a dense breast shown in Fig. 8.9, with a volume fraction of dense tissue of 28 %, a reliable estimate could be obtained with our method.

To our knowledge, no other studies have been published that investigate volumetric breast density estimation from FFD mammograms. In previous studies, digitized mammograms were used and breast density was most often measured in qualitative categories or as the size of the projected dense tissue area. Our work resembles the approach proposed by Highnam & Brady (1999), who proposed a standardized mammogram form in which pixel values represent dense tissue thickness. Application of their method is complicated, however, as it takes many aspects of the image formation process into account in a physics based model. Consequently, it requires many parameters related to the acquisition procedure to be known. A major difference between our method and SMF

is the internal calibration we use by obtaining a fatty tissue reference. In this way, our method gets less dependent on breast thickness, for which accurate measurements are hard to obtain in practice. Without internal calibration, small errors in breast thickness have a strong effect on density estimates. The SMF approach has been used for volumetric breast density estimation in digitized mammograms in a study by Marias *et al.* (2004). In their work performance was evaluated by assessing agreement with the opinion of a clinician, without further quantitative validation.

Several sources of error can be identified to explain fluctuations in Fig. 8.5 and 8.6 apart from inaccuracy of the ground truth. As mentioned above, one source of error lies in the calibration measurement, in particular when the breast is very dense. Other sources of errors are due to neglecting scattered radiation and to errors in the attenuation coefficients. With respect to the effect of scatter it is remarked that it is common to use anti-scatter grids in mammography, which effectively decrease influence of scatter. It will be worthwhile though to study the effect of the remaining scatter on our method in detail. The same holds for the effect of patient variability of the attenuation coefficients.

Our method uses two tissue types, namely fatty and dense tissue. One might argue that the skin, with attenuation coefficient closer to dense than to fatty tissue influences our results. However, the contribution due to the skin will also be present in the fatty tissue reference value. Therefore, the attenuation term related to the skin will cancel out in Eq. 8.7, when assuming that the skin thickness is constant over the breast area.

In this study we used MRI data and an interactive segmentation method to establish a ground truth measure for fibro-glandular tissue volume. This approach has been used before by Stoutjesdijk & Karssemeijer (2000) and Wei *et al.* (2004), who both determined the projected dense tissue area from digitized mammograms and found a good correlation between the mammogram area density and the volumetric density from MRI data. It is likely, however, that breast cancer risk is more strongly related to the volume of dense tissue than to the projected area of dense tissue on a mammogram. Our study suggests that by using FFD mammography accurate estimation of dense tissue volume becomes feasible.

Bibliography

BASSETT, L.W., D.H. BUNNELL, R. JAHANSHAHI, R.H. GOLD, R.D. ARNDT, & J. LINSMAN. 1987. Breast cancer detection: one versus two views. *Radiology* 165:95–97.

——, B. SHAYESTEHFAR, & I. HIRBAWI. 1994. Obtaining previous mammograms for comparison: usefulness and costs. *American Journal of Roentgenology* 163:1083–1086.

BICK, U., M.L. GIGER, R.A. SCHMIDT, R.M. NISHIKAWA, & K. DOI. 1996. Density correction of peripheral breast tissue on digital mammograms. *Radiographics* 16(6):1403–1411.

BIRD, R.E., T.W. WALLACE, & B.C. YANKASKAS. 1992. Analysis of cancers missed at screening mammography. *Radiology* 184:613–617.

BLANKS, R.G., M.G. WALLIS, & R.M. GIVEN-WILSON. 1999. Observer variability in cancer detection during routine repeat (incident) mammographic screening in a study of two versus one view mammography. *Journal of Medical Screening* 6:152–158.

——, M.G. WALLIS, & S.M. MOSS. 1998. A comparison of cancer detection rates achieved by breast cancer screening programmes by number of readers, for one and two view mammography: results from the UK national health service breast screening programme. *Journal of Medical Screening* 5:195–201.

BOOKSTEIN, F.L. 1989. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11(6):567–585.

BOONE, J.M., T.R. FEWELL, & R.J. JENNINGS. 1997. Molybdenum, rhodium, and tungsten anode spectral models using interpolating polynomials with application to mammography. *Medical Physics* 24(12):1863–1874.

BORGEFORS, G. 1986. Distance transformation in digital images. *Computer Vision, Graphics and Image Processing* 34:344–371.

BURGESS, A.E. 2004. On the noise variance of a digital mammography system. *Medical Physics* 31(7):1987–1995.

——, F.J. JACOBSON, & P.F. JUDY. 2001. Human observer detection experiments with mammograms and power-law noise. *Medical Physics* 28(4):419–437.

BURHENNE, L.J.W., S.A. WOOD, C.J. D'ORSI, S.A. FEIG, D.B. KOPANS, K.F. O'SHAUGHNESSY, E.A. SICKLES, L. TABAR, C.J. VYBORNY, & R.A. CASTELLINO. 2000. Potential contribution of computer-aided detection to sensitivity of screening mammography. *Radiology* 215:554–562.

BURNSIDE, E.S., E.A. SICKLES, R.E. SOHLICH, & K.E. DEE. 2002. Differential value of comparison with previous examinations in diagnostic versus screening mammography. *American Journal of Roentgenology* 179:1173–1177.

BYNG, J.W., N.F. BOYD, E. FISHELL, R.A. JONG, & M.J. YAFFE. 1994. The quantitative analysis of mammographic densities. *Physics in Medicine and Biology* 39:1629–1638.

——, J.P. CRITTEN, & M.J. YAFFE. 1997. Thickness-equalization processing for mammographic images. *Radiology* 203:564–568.

——, J.G. MAINPRIZE, & M.J. YAFFE. 1998a. X-ray characterization of breast phantom materials. *Phys. Med. Biol.* 43:1367–1377.

——, M.J. YAFFE, R.A. JONG, R.S. SHUMAK, G.A. LOCKWOOD, D.L. TRITCHLER, & N.F. BOYD. 1998b. Analysis of mammographic density and breast cancer risk from digitized mammograms. *Radiographics* 18(6):1587–1598.

BYRNE, C., C. SCHAIRER, J. WOLFE, N. PAREKH, M. SALANE, L.A. BRINTON, R. HOOVER, & R. HAILE. 1995. Mammographic features and breast cancer risk: effects with time, age, and menopause status. *Journal of the National Cancer Institute* 87(21):1622–1629.

CALDWELL, C.B., S.J. STAPLETON, D.W. HOLDSWORTH, R.A. JONG, W.J. WEISER, G. COOKE, & M.J. YAFFE. 1990. Characterisation of mammographic parenchymal pattern by fractal dimension. *Physics in Medicine and Biology* 35(2):235–247.

CALLAWAY, M.P., C.R.M. BOGGIS, S.A. ASTLEY, & I. HUTT. 1997. The influence of previous films on screening mammographic interpretation and detection of breast carcinoma. *Clinical Radiology* 52:527–529.

CHAN, H., B. SAHINER, M. HELVIE, N. PETRICK, M. ROUBIDOUX, T. WILSON, D. ADLER, C. PARAMAGUL, J. NEWMAN, & S. SANJAY-GOPAL. 1999. Improvement of radiologists' characterization of mammographic masses by using computer-aided diagnosis: An ROC study. *Radiology* 212:817–827.

COLLIGNON, A., F. MAES, D. DELAERE, D. VANDERMEULEN, P. SUETENS, & G. MARCHAL. 1995. Automated multimodality medical image registration using information theory. In *Information Processing in Medical Imaging; Computational Imaging and Vision 3*, ed. by Y. Bizais, C. Barillot, & R. Di Piola, pages 263–274. Kluwer.

CUPPLES, T.E., J.E. CUNNINGHAM, & J.C. REYNOLDS. 2005. Impact of computer-aided detection in a regional screening mammography program. *American Journal of Roentgenology* 185(4):944–950.

DORFMAN, D.D., K.S. BERBAUM, & C.E. METZ. 1992. ROC rating analysis: generalization to the population of readers and cases with the jackknife method. *Investigative Radiology* 27:723–731.

DRONKERS, D.J., J.H.C.L. HENDRIKS, R. HOLLAND, & G. ROSENBUSCH. 2002. *The practice of mammography*. Thieme New York.

FEIG, S.A., E.A. SICKLES, W.P. EVANS, & M.N. LINVER. 2004. Re: Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system. *Journal of the National Cancer Institute* 96(16):1260–1261.

FILEV, P., L. HADJIISKI, B. SAHINER, H. CHAN, & M.A. HELVIE. 2005. Comparison of similarity measures for the task of template matching of masses on serial mammograms. *Medical Physics* 32(2):515–29.

FRANKEL, S.D., E.A. SICKLES, B.N. CURPEN, R.A. SOLLITTO, S.H. OMINSKY, & H.B. GALVIN. 1995. Initial versus subsequent screening mammography: comparison of findings and their prognostic significance. *American Journal of Roentgenology* 164:1107–1109.

FREER, T., & M. ULISSEY. 2001. Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center. *Radiology* 220:781–786.

FUTREAL, P.A., Q. LIU, D. SHATTUCK-EIDENS, C. COCHRAN, K. HARSHMAN, S. TAVTIGIAN, L.M. BENNETT, A. HAUGEN-STRANO, J. SWENSEN, & Y. MIKI. 1994. BRCA1 mutations in primary breast and ovarian carcinomas. *Science* 266(5182):120–122.

GOOD, W., B. ZHENG, Y. CHANG, X. WANG, G. MAITZ, & D. GUR. 1999. Multi-image CAD employing features derived from ipsilateral mammographic views. In *Medical Imaging: Image Processing*, ed. by K.M. Hanson, volume 3661, pages 474–485. Proceedings of SPIE.

GUR, D., D. CHOUGH, C. COHEN, J.H. SUMKIN, G. ABRAMS, M.A. GANOTT, L. WALLACE, R. SHAH, & B. ZHENG. 2006. Actual versus intended use of CAD systems in the clinical environment. In *Medical Imaging: Image Perception, Observer Performance, and Technology Assessment*, volume 6146, page to appear. Proceedings of SPIE.

——, J.H. SUMKIN, H.E. ROCKETTE, M. GANOTT, C. HAKIM, L. HARDESTY, W.R. POLLER, R. SHAH, & L. WALLACE. 2004. Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system. *Journal of the National Cancer Institute* 96(3):185–190.

HADJIISKI, L., H. CHAN, B. SAHINER, N. PETRICK, & M.A. HELVIE. 2001a. Automated registration of breast lesions in temporal pairs of mammograms for interval change analysis - local affine transformation for improved localization. *Medical Physics* 28(6):1070–1079.

——, B. SAHINER, H. CHAN, N. PETRICK, M.A. HELVIE, & M. GURCAN. 2001b. Analysis of temporal changes of mammographic features: Computer-aided classification of malignant and benign breast masses. *Medical Physics* 28:2309–2317.

HARVEY, J.A., & V.E. BOVBJERG. 2004. Quantitative assessment of mammographic breast density: Relationship with breast cancer risk. *Radiology* 230:29–41.

HEINE, J.J., & P. MALHOTRA. 2002a. Mammographic tissue, breast cancer risk, serial image analysis, and digital mammography. part 1. tissue and related risk factors. *Academic Radiology* 9(3):298–316.

——, & ——. 2002b. Mammographic tissue, breast cancer risk, serial image analysis, and digital mammography. part 2. serial breast tissue change and related temporal influences. *Academic Radiology* 9(3):317–335.

HIGHNAM, R., Y. KITA, J.M. BRADY, B. SHEPSTONE, & R. ENGLISH. 1998. Determining correspondence between views. In *Digital Mammography*, pages 111–118, Nijmegen. Kluwer Academic Publishers.

HIGHNAM, R.P., & J.M. BRADY. 1999. *Mammographic Image Analysis*. Dordrecht: Kluwer.

HUBBELL, J.H., & S.M. SELTZER. 1995. Tables of X-Ray mass attenuation coefficients and mass energy-absorption coefficients. from $1\,keV$ to $20\,MeV$ for elements $Z = 1$ to 92 and 48 additional substances of dosimetric interest. Technical report, National Institute of Standards and Technology, report NISTIR 5632.

HUO, Z., M.L. GIGER, O.I. OLOPADE, D.E. WOLVERTON, B.L. WEBER, C.E. METZ, W. ZHONG, & S.A. CUMMINGS. 2002. Computerized analysis of digitized mammograms of BRCA1 and BRCA2 gene mutation carriers. *Radiology* 225(2):519–526.

——, M.L. GIGER, D.E. WOLVERTON, W. ZHONG, S. CUMMING, & O.I. OLOPADE. 2000. Computerized analysis of mammographic parenchymal patterns for breast cancer risk assessment: feature selection. *Medical Physics* 27(1):4–12.

JIANG, Y.L., R.M. NISHIKAWA, C.E. METZ R.A. SCHMIDT, M.L. GIGER, & K. DOI. 1999. Improving breast cancer diagnosis with computer-aided diagnosis. *Academic Radiology* 6:22–33.

JOHNS, P.C., & M.J. YAFFE. 1987. X-ray characterization of normal and neoplastic tissues. *Physics in Medicine and Biology* 32:675–695.

JONKER, R., & A. VOLGENANT. 1987. A shortest augmenting path algorithm for dense and sparse linear assignment problems. *Computing* 38:325–340.

KARSSEMEIJER, N. 1998. Automated classification of parenchymal patterns in mammograms. *Physics in Medicine and Biology* 43:365–378.

——, J.D.M. OTTEN, A.L.M. VERBEEK, J.H. GROENEWOUD, H.J. DE KONING, J.H.C.L. HENDRIKS, & R. HOLLAND. 2003. Computer-aided detection versus independent double reading of masses on mammograms. *Radiology* 227:192–200.

——, & G.M. TE BRAKE. 1996. Detection of stellate distortions in mammograms. *IEEE Transactions on Medical Imaging* 15:611–619.

——, & ——. 1998. Combining single view features and asymmetry for detection of mass lesions. In *Digital Mammography*, ed. by N. Karssemeijer, M.A.O. Thijssen, J.H.C.L. Hendriks, & L.J.T.O. van Erning, pages 95–102. Kluwer, Dordrecht.

KRIEGE, M., C.T. BREKELMANS, C. BOETES, E.J. RUTGERS, J.C. OOSTERWIJK, R.A. TOLLENAAR, R.A. MANOLIU, R. HOLLAND, H.J. DE KONING, & J.G. KLIJN. 2001. MRI screening for breast cancer in women with familial or genetic predisposition; design of the dutch national study (MRISC). *Familial Cancer* 1(3-4):163–168.

KUNDEL, H.L. 2000. *Visual search in medical images*, volume 1. Physics and Psychophysics, chapter 18.2.5, pages 837–858. Bellingham, WA SPIE.

LAU, T., & W. BISCHOF. 1991. Automated detection of breast tumors using the asymmetry approach. *Computers and Biomedical Research* 24:273–295.

MARIAS, K., C. BEHRENBRUCH, R. HIGHNAM, S. PARBHOO, A. SEIFALIAN, & J.M. BRADY. 2004. A mammographic image analysis method to detect and measure changes in breast density. *European Journal Radiology* 52:276–282.

METZ, C.E. 1986. ROC methodology in radiographic imaging. *Investigative Radiology* 21:720–733.

NISHIKAWA, R.M., A. EDWARDS, R.A. SCHMIDT, J. PAPAIOANNOU, & M.N. LINVER. 2006. Can radiologists recognize that a computer has identified cancers that they have overlooked? In *Medical Imaging: Image Perception, Observer Performance, and Technology Assessment*, volume 6146, page to appear. Proceedings of SPIE.

OTTEN, J.D.M., N. KARSSEMEIJER, J.H.C.L. HENDRIKS, J.H. GROENEWOUD, J. FRACHEBOUD, A.L.M. VERBEEK, H.J. DE KONING, & R. HOLLAND. 2005. Effect of recall rate on earlier screen detection of breast cancers based on the dutch performance indicators. *Journal of the National Cancer Institute* 10:748–754.

OTTO, S.J., J. FRACHEBOUD, C.W.N. LOOMAN, M.J.M. BROEDERS, R. BOER, J.H.C.L. HENDRIKS, A.L.M. VERBEEK, & H.J. DE KONING. 2003. Initiation of population-based mammography screening in Dutch municipalities and effect on breast-cancer mortality: a systematic review. *Lancet* 361(9367):1411–1417.

PAQUERAULT, S., N. PETRICK, H. CHAN, B. SAHINER, & M.A. HELVIE. 2002. Improvement of computerized mass detection on mammograms: Fusion of two-view information. *Medical Physics* 29(2):238–247.

PAWLUCZYK, O., B. J. AUGUSTINE, M. J. YAFFE, D. RICO, J. YANG, G. E. MAWDSLEY, & N.F. BOYD. 2003. A volumetric method for estimation of breast density on digitized screen-film mammograms. *Medical Physics* 30(3):352–364.

PRESS, W., B. FLANNERY, S. TEUKOLSKY, & W. VETTERLING. 1992. *Numerical Recipes in C*. Cambridge University Press, 2nd edition.

SAHA, P.K., J.K. UDUPA, E.F. CONANT, D.P. CHAKRABORTY, & D. SULLIVAN. 2001. Breast tissue density quantification via digitized mammograms. *IEEE Transactions on Medical Imaging* 20(8):792–803.

SAJDA, P., C. SPENCE, & J. PEARSON. 2002. Learning contextual relationships in mammograms using a hierarchical pyramid neural network. *IEEE Transactions on Medical Imaging* 21(3):239–250.

SALLAM, M., & K.W. BOWYER. 1996. Registering time-sequences of mammograms using a two-dimensional unwarping technique. In *Digital Mammography*, ed. by K. Doi, M.L. Giger, R.M. Nishikawa, & R.A. Schmidt, pages 291–296. Elsevier, Amsterdam.

SANJAY-GOPAL, S., H. CHAN, T. WILSON, M.A. HELVIE, N. PETRICK, & B. SAHINER. 1999. A regional registration technique for automated interval change analysis of breast lesions on mammograms. *Medical Physics* 26(12):2669–2679.

SICKLES, E.A. 1991. Periodic mammographic follow-up of probably benign lesions: results in 3,184 consecutive cases. *Radiology* 179:463–468.

SNOEREN, P.R., & N. KARSSEMEIJER. 2004. Thickness correction of mammographic images by means of a global parameter model of the compressed breast. *IEEE Transactions on Medical Imaging* 23(7):799–806.

——, & ——. 2005. Thickness correction of mammographic images by anisotropic filtering and interpolation of dense tissue. In *Medical Imaging: Image Processing*, ed. by J.M. Fitzpatrick & J.M. Reinhardt, volume 5747, pages 1521–1527. Proceedings of SPIE.

——, & M.J.H. PUTS. 1997. Multiple parameter estimation in an adaptive psychometric method: MUEST, an extension of the QUEST method. *Journal of Mathematical Psychology* 41(4):431–439.

STOUTJESDIJK, M.J., & N. KARSSEMEIJER. 2000. Validation of a method for segmentation of dense tissue in mammograms by correlation with MRI data. In *Digital Mammography*, ed. by M. J. Yaffe, pages 770–776, Toronto, Canada. Medical Physics Publishing.

SWENSSON, R.G., S.J. HESSEL, & P. HERMAN. 1982. Radiographic interpretation with and without search: Visual search aids the recognition of chest pathology. *Investigative Radiology* 17(2):145–151.

TAHOCES, P.G., J. CORREA, M. SOUTOS, L. GOMEZ, & J.J. VIDAL. 1995. Computer-assisted diagnosis: the classification of mammographic breast parenchymal patterns. *Physics in Medicine and Biology* 40(1):103–117.

TE BRAKE, G.M., & N. KARSSEMEIJER. 1999. Single and multiscale detection of masses in digital mammograms. *IEEE Transactions on Medical Imaging* 18(7):628–639.

THURFJELL, E. 1994. Mammography screening: One versus two views and independent double reading. *Acta Radiologica* 35:345–350.

THURFJELL, M.G., B. VITAK, E. AZAVEDO, G. SVANE, & E. THURFJELL. 2000. Effect on sensitivity and specificity of mammography screening with or without comparison of old mammograms. *Acta Radiologica* 41:52–56.

TIMP, S., & N. KARSSEMEIJER. 2004. A new 2D segmentation method based on dynamic programming applied to computer aided detection in mammography. *Medical Physics* 31(5):958–971.

——, & ——. 2006. Interval change analysis to improve computer aided detection in mammography. *Medical Image Analysis* 10(1):82–95.

——, S. VAN ENGELAND, & N. KARSSEMEIJER. 2005. A regional registration method to find corresponding mass lesions in temporal mammogram pairs. *Medical Physics* 32(8):2629–2638.

VAN ENGELAND, S., & N. KARSSEMEIJER. 2001. Matching breast lesions in multiple mammographic views. In *Medical Image Computing and Computer-Assisted Intervention*, ed. by W Niessen & M Viergever, volume 2208, pages 1172–1173. Springer, LNCS.

——, & ——. 2005. Regrouping initial CAD mass detections to facilitate classification of suspicious regions in mammography. In *Medical Imaging: Image Processing*, ed. by J.M. Fitzpatrick & J.M. Reinhardt, volume 5747, pages 975–986. Proceedings of SPIE.

——, ——, & J.H.C.L. HENDRIKS. 2002. Using information from two mammographic views to improve computer-aided detection of mass lesions. In *6th International Workshop on Digital Mammography*, ed. by H. Peitgen, pages 377–381. Springer-Verlag, Berlin.

——, P.R. SNOEREN, J.H.C.L. HENDRIKS, & N. KARSSEMEIJER. 2003a. A comparison of methods for mammogram registration. *IEEE Transactions on Medical Imaging* 22(11):1436–1444.

——, P.R. SNOEREN, H. HUISMAN, C. BOETES, & N. KARSSEMEIJER. 2006. Volumetric breast density estimation from full field digital mammograms. *IEEE Transactions on Medical Imaging* 3(25):273–282.

——, P.R. SNOEREN, N. KARSSEMEIJER, & J.H.C.L. HENDRIKS. 2003b. Optimized perception of lesion growth in mammograms using digital display. In *Medical Imaging: Image Perception, Observer Performance, and Technology Assessment*, ed. by D. Chakraborty & E. Krupinski, volume 5034, pages 25–31. Proceedings of SPIE.

——, C. VARELA, S. TIMP, P.R. SNOEREN, & N. KARSSEMEIJER. 2005. Using context for mass detection and classification in mammograms. In *Medical Imaging: Image*

Perception, Observer Performance, and Technology Assessment, ed. by M.P. Eckstein & Y. Jiang, pages 94–102. Proceedings of SPIE.

VAN GILS, C.H., J.H. HENDRIKS, R. HOLLAND, N. KARSSEMEIJER, J.D. OTTEN, H. STRAATMAN, & A.L. VERBEEK. 1999a. Changes in mammographic breast density and concomitant changes in breast cancer risk. *European Journal of Cancer Prevention* 8(6):509–515.

——, J.D. OTTEN, J.H. HENDRIKS, R. HOLLAND, H. STRAATMAN, & A.L. VERBEEK. 1999b. High mammographic breast density and its implications for the early detection of breast cancer. *Journal of Medical Screening* 6(4):200–204.

VELDKAMP, W.J.H., N. KARSSEMEIJER, J.D. OTTEN, & J.H.C.L. HENDRIKS. 2000. Automated classification of clustered microcalcifications into malignant and benign types. *Medical Physics* 27:2600–2608.

VIOLA, P., & W. WELLS. 1995. Alignment by maximization of mutual information. In *International Conference on Computer Vision*, ed. by E. Grimson, S. Shafer, A. Blake, & K. Sugihara, pages 16–23. Los Alamitos, CA, IEEE Computer Society Press.

VUJOVIC, N., & D. BRZAKOVIC. 1997. Establishing the correspondence between control points in pairs of mammographic images. *IEEE Transactions on Medical Imaging* 6(10):1388–1399.

WANG, X.H., B.E. CHAPMAN, C.A. BRITTON, S.K. GOLLA, L.P. WALLACE, & W.F. GOOD. 2003a. Evaluation of quantitative measures of breast tissue density from mammography with truth from MRI data. In *Medical Imaging: Image Perception, Observer Performance, and Technology Assessment*, ed. by D. Chakraborty & E. Krupinski, volume 5034, pages 82–88. Proceedings of SPIE.

——, W.F. GOOD, B.E. CHAPMAN, Y.H. CHANG, W.R. POLLER, T.S. CHANG, & L.A. HARDESTY. 2003b. Automated assessment of the composition of breast tissue revealed on tissue-thickness-corrected mammography. *American Journal of Roentgenology* 180(1):257–262.

WEI, J., H. CHAN, M.A. HELVIE, M.A. ROUBIDOUX, B. SAHINER, L.M. HADJIISKI, C. ZHOU, S. PAQUERAULT, T. CHENEVERT, & M.M. GOODSITT. 2004. Correlation between mammographic density and volumetric fibroglandular tissue estimated on MR images. *Medical Physics* 31(4):933–942.

WILSON, T.E., V.K. NIJHAWAN, & M.A. HELVIE. 1996. Normal mammograms and the practice of obtaining previous mammograms: usefulness and costs. *Radiology* 198:661–663.

WIRTH, M.A., J. NARHAN, & D. GRAY. 2002. Nonrigid mammogram registration using mutual information. In *Medical Imaging: Image Processing*, ed. by M. Sonka & J.M. Fitzpatrick, volume 4684, pages 562–573. Proceedings of SPIE.

WIVELL, G., E.R. DENTON, J.C. INGLIS, & I. HARVEY. 2003. Can radiographers read screening mammograms? *Clinical Radiology* 58(1):63–67.

WOLFE, J.N. 1976. Risk for breast cancer development determined by mammographic parenchymal pattern. *Cancer* pages 2486–2492.

YIN, F., M.L. GIGER, K. DOI, C.J. VYBORNY, & R.A. SCHMIDT. 1994. Computerized detection of masses in digital mammograms: Automated alignment of breast images and its effect on bilateral-subtraction technique. *Medical Physics* 21(3):445–452.

ZHOU, C., H. CHAN, N. PETRICK, M.A. HELVIE, M.M. GOODSITT, B. SAHINER, & L. HADJIISKI. 2001. Computerized image analysis: estimation of breast density on mammograms. *Medical Physics* 28(6):1056–1069.

Summary

Computer-aided Detection (CAD) and image processing techniques for mammography are being developed to aid radiologists during screening and to increase the detection rate. During interpretation, radiologists use information from all available views, i.e. previous examinations and projections from different angles. Most current CAD systems, on the other hand, use information from only one view at the time. The research described in this thesis concerns the development of multiple view CAD techniques to increase the performance of CAD for the detection of masses in mammograms. The first part of this thesis addresses the comparison of temporal mammogram pairs. The second part concerns the combination of information from two projections of the same breast. The final chapter describes a method for accurate quantitative estimation of the dense tissue volume from mammograms. Such a measure for breast density can be used for instance for comparisons of the left and right breast.

Mammogram registration is an important technique to optimize the display of cases on a digital viewing station. In order to compare two mammographic views they need to be correctly registered, ensuring that a (possible) lesion is displayed at about the same location on the screen for both views. It was investigated which method is the most appropriate for the registration of temporal mammogram pairs. The performance of four registration methods was measured by comparing the distance between annotations of abnormalities in the previous and current view before and after registration. Registration by mutual information outperformed alignment based on nipple location, alignment based on center of mass of breast tissue and warping. In addition to being useful for displaying, a mammogram registration technique can also be used for the development of CAD algorithms that use temporal information.

Growth of lesions is an important clue to detect lesions and to discriminate between benign and malignant abnormalities. It was investigated whether it is possible to improve the detection of lesion growth by using image processing. Two ways of presenting prior and current mammograms on a mammography workstation were compared: (1) display next to each other and (2) alternating at the same display (toggle). In an observer experiment, 420 trials with prior-current mammogram pairs were displayed on a dedicated

mammography workstation. In a two-alternative forced-choice experiment, observers were asked to select the image containing the largest lesion. The stimuli were created by pasting extracted lesions into normal mammograms. Results showed that the observers performed more accurate in selecting the largest lesion when using the toggle option.

Separate parts of relatively large lesions are sometimes detected as individual lesions by the CAD software. The development of multiple view techniques is hampered by this phenomena. To facilitate the final classification of suspicious regions, a technique to regroup initial candidate regions of our CAD scheme was developed. This regrouping technique identifies candidate regions that belong to the same structure, taking into account the distance between the regions and the image structure along a path between the regions. When correspondence is found, the two regions are replaced by a new region in between the initial candidate regions. The regrouping technique correctly regrouped the candidate regions in 48 percent of the masses initially detected by multiple regions. Of the false positive regions two percent were combined, and the percentage of true positive - false positive combinations was one.

During screening, the mediolateral oblique (MLO) and the cranio-caudal (CC) views are the most often used projections. For the development of CAD algorithms that use combined information from these views, corresponding regions in both views need to be found. To this end, we have developed a method in which for all possible combinations of candidate regions, features are calculated. These features describe the difference in the radial distance from the regions to the nipple, gray scale correlation between the two regions and the 'mass likelihood' of both regions. Linear Discriminant Analysis is used to discriminate between correct and incorrect correspondences. The method was tested on an annotated set of 412 cases. Results showed that in 82 % of the image pairs a correct link between the true positive regions in both views was established.

Using the established correspondences between regions in MLO and CC views, a technique was developed to improve mass detection results. To this end, our CAD scheme was extended by including another classifier that uses two view features. These two view features describe the resemblance of two corresponding regions. The performance of the two view classifier was compared with the detection performance of our original CAD scheme using Free-response Receiver Operating Characteristic analysis. The image based evaluation showed an improvement of 15 % in sensitivity at a false positive rate of 0.1 FP/image.

Most CAD schemes use local features. There is a lot of interest in developing CAD methods that use context, asymmetry, and multiple view information. It was expected that the use of this extra information would improve CAD results. Concluding from studies described in literature so far, however, the improvements of CAD results are either small or not clear. Therefore, it was investigated to what extend human readers make use of context information derived from the whole breast area and from asymmetry for the tasks of mass detection. The results from this study showed that context information

can be used to improve CAD programs for mass detection. However, there is still a lot to be gained from improvement of local feature extraction and classification. This was demonstrated by the fact that the observers did much better in classifying true positive and false positive regions than the CAD program, even when only local information was presented to the reader. These results can guide further development of CAD schemes.

The final chapter of this thesis describes a method for the estimation of dense breast tissue volume from mammograms obtained with full field digital mammography. Automatic determination of breast tissue density is important since breast tissue density has been identified as an important risk factor for breast cancer development. In addition, such a measure for breast density might be used for instance for the comparison of the left and the right breast, and for CAD methods that use context information. The thickness of dense tissue mapping to a pixel was determined by using a physical model of image acquisition. This model is based on the assumption that the breast is composed of two types of tissue, fat and parenchyma. Effective linear attenuation coefficients of these tissues were derived from empirical data as a function of tube voltage, anode material, filtration, and compressed breast thickness. By employing these, tissue composition at a given pixel was computed after performing breast thickness compensation, using a reference value for fatty tissue determined by the maximum pixel value in the breast tissue projection. Validation was performed using 22 cases by comparing the volume estimates with volumes obtained by semi-automatic segmentation of breast Magnetic Resonance Imaging (MRI) data. The correlation between MRI and mammography volumes was high, 0.94 on a per image basis and 0.97 on a per patient basis.

Samenvatting

Voor de vroege opsporing van borstkanker zijn er in veel landen bevolkingsonderzoeken opgestart (borstkankerscreening). Hierbij worden vrouwen in een bepaalde leeftijdscategorie periodiek uitgenodigd om röntgenfoto's van beide borsten (mammogram) te laten maken. Om radiologen tijdens de beoordeling deze mammogrammen te ondersteunen bij de opsporing van tumoren zijn nieuwe methoden in ontwikkeling, zogenaamde Computer-aided Detection (CAD) methoden. Hierbij wordt de computer gebruikt om met beeldherkenningstechnieken verdachte gebieden in mammogrammen aan te wijzen. Tot nu toe zijn vooral methoden ontwikkeld waarbij de computer slechts naar een enkele opname van de borst 'kijkt'. Hiermee is het mogelijk om het grootste deel van de tumoren correct aan te wijzen. In de praktijk maken radiologen echter gebruik van veel meer informatie. Zo vergelijken zij opnames van de linker en de rechter borst, opnames van dezelfde borst gemaakt onder verschillende hoeken, en opnames die tijdens eerdere screeningsronden gemaakt zijn.

Het werk dat in dit proefschrift beschreven is betreft de ontwikkeling van computer detectiemethoden die meer overeenkomen met de beoordelingswijze van radiologen. Het uiteindelijke doel hierbij is het verminderen van het aantal gemiste tumoren en het verminderen van het aantal onterecht (fout positief) als verdacht aangewezen gebieden. De eerste twee hoofdstukken hebben betrekking op het vergelijken met mammogrammen die tijdens een eerdere screeningsronde gemaakt zijn. Daarna volgen vier hoofdstukken over het combineren van twee verschillende opnames van dezelfde borst gemaakt tijdens dezelfde screeningsronde. Het laatste hoofdstuk beschrijft een methode voor de automatische bepaling van de klierweefseldichtheid uit een mammogram. Dit kan bijvoorbeeld gebruikt worden bij het vergelijken van de linker en de rechter borst.

Voor het presenteren van digitale mammogrammen is het van belang dat mammogrammen goed zijn geregistreerd. Dit zorgt ervoor dat twee verschillende opnames van dezelfde borst beter met elkaar vergeleken kunnen worden. Twee opnames zijn goed geregistreerd als de (mogelijke) tumor voor zowel de voorgaande als de huidige opname op ongeveer dezelfde positie op het beeldscherm wordt weergegeven. Om de voorgaande opname zo goed mogelijk op de huidige opname te doen lijken wat betreft positionering,

kan de voorgaande opname bijvoorbeeld getransleerd, geroteerd en geschaald worden. Er zijn vier methoden voor het registreren van mammogrammmen van de voorgaande en de huidige screeningsronde (temporele mammogram paren) getest. Voor de evaluatie zijn de afstanden tussen de tumoren op de voorgaande opname en de huidige opname voor en na registratie vergeleken. Registratie met 'mutual information', een registratiemaat die gebaseerd is op grijswaarde combinaties in twee beelden, gaf de beste resultaten. De andere methoden die getest zijn waren; uitlijnen op basis van de tepel locatie, uitlijnen op basis van het zwaartepunt van de borst en een methode waarbij op basis van punten op de borstcontour een 'mapping' is gemaakt van de voorgaande opname naar de huidige. Het registreren van mammogrammen kan ook gebruikt worden bij het ontwikkelen van CAD algoritmen waarbij temporele informatie gebruikt wordt.

Groei van een verdichting op een mammogram is een belangrijke aanwijzing bij het opsporen van borstkanker. Daarnaast is het herkennen van groei van belang bij het maken van onderscheid tussen goedaardige en kwaadaardige afwijkingen. Twee temporele opnames kunnen op twee manieren gepresenteerd worden, naast elkaar en achter elkaar (toggle-mode) op het scherm. Beide methoden zijn vergeleken met het oog op de herkenbaarheid van groei van verdichtingen. Tijdens een perceptie-experiment met vijf waarnemers is 420 keer een afwijking gepresenteerd in opnames van twee op elkaar volgende tijdsmomenten. Tijdens het experiment is de waarnemers gevraagd een keuze maken in welk plaatje de grootste verdichtingen te zien was. Het bleek dat de waarnemers beter waren in het aanwijzen van de grootste verdichting wanneer de toggle-mode werd gebruikt.

Soms gebeurt het dat afzonderlijke delen van grote tumoren als losse tumoren herkend worden door de CAD software. Dit verschijnsel bemoeilijkt de ontwikkeling van detectietechnieken waarbij informatie uit meerdere opnames tegelijk gebruikt wordt. Om de uiteindelijke classificatie van verdachte gebieden makkelijker te maken, is een techniek ontwikkeld die verdachte kandidaatgebieden hergroepeert. Deze hergroeperingsmethode zoekt naar kandidaatgebieden die tot dezelfde structuur behoorden. Daarvoor is gekeken naar de afstand tussen de kandidaatgebieden en de beeldstructuur langs een pad tussen de twee gebieden. Wanneer wordt gevonden dat beide gebieden tot dezelfde structuur behoorden, worden deze vervangen door een nieuwe kandidaatgebied tussen de oorspronkelijke gebieden in. Met deze methode is het gelukt om in 48 procent van de tumoren, die initieel gedetecteerd werden door meerdere kandidaatgebieden, de gebieden succesvol te hergroeperen. Van de door het CAD programma fout positief aangemerkte gebieden is twee procent gerecombineerd en het percentage terecht positieve - fout positieve combinaties was een.

De twee opnamerichtingen gebruikt bij de screening zijn de mediolateral oblique (MLO) en cranio-caudal (CC) opnames. Voor CAD algoritmen die gecombineerde informatie uit deze opnames gebruiken is het nodig overeenkomstige gebieden in beide projecties te vinden. Hiervoor is een methode ontwikkeld, waarbij voor alle mogeli-

jke combinaties van verdachte gebieden uit beide opnames, is een aantal features (kenmerken) berekend. Deze features hebben betrekking op het verschil in afstand van het gebied tot de tepel, de grijswaarde correlatie tussen de twee gebieden en de 'tumorwaarschijnlijkheid' van beide gebieden. Lineaire Discriminanten Analyse is gebruikt om onderscheid te maken tussen correcte en incorrecte combinaties. Deze methode is getest op een set van 412 cases. De resultaten laten zien dat in 82 % van de MLO/CC opname paren een terecht positief gebied in de ene opname correct werd gekoppeld aan een terecht positief gebied in de andere opname.

Door gebruik te maken van corresponderende gebieden in MLO en CC opnames is een techniek ontwikkeld om de tumordetectie verder te verbeteren. Daartoe is het CAD programma uitgebreid met een derde 'classifier', een automatische leertechniek afkomstig uit de kunstmatige intelligentie. De invoer voor deze classifier zijn combinatie features, kenmerken die afhangen van de mate waarin de gekoppelde gebieden op elkaar lijken. De resultaten van deze classifier zijn vergeleken met die van het oorspronkelijke CAD programma met behulp van 'Free-response Receiver Operating Characteristic' analyse, waarbij een afweging wordt gemaakt van de detectie-waarschijnlijkheid tegen het aantal onterecht aangegeven fout positieve gebieden. Voor 0.1 fout positief per opname leidt dit tot een verbetering van 15 % in sensitiviteit.

De meeste CAD software gebruikt lokaal bepaalde beeldkenmerken. Er is veel belangstelling voor het ontwikkelen van CAD methoden die contextinformatie, asymmetrie en de informatie uit meerdere opnames tegelijk gebruiken. Echter uit meerdere studies in de literatuur blijkt dat deze extra informatie niet leidt tot een duidelijke verbetering van de detectie resultaten. Daarom is er onderzocht in welke mate menselijke waarnemers gebruik maken van contextinformatie afgeleid uit de hele opname van de borst en asymmetrie voor tumordetectie. De resultaten laten zien dat contextinformatie de waarnemers inderdaad helpt bij het detecteren van tumoren. Het is echter ook gebleken dat er voor CAD programma's nog veel winst valt te behalen uit de verbetering van lokale feature extractie. Dit is afgeleid uit het feit dat de waarnemers duidelijk beter waren in het classificeren van gebieden in echte abnormaliteiten en fout positieven dan het CAD programma. Deze resultaten kunnen gebruikt worden voor het sturen van de verdere ontwikkeling van CAD software.

Het laatste hoofdstuk van dit proefschrift gaat over de bepaling van de hoeveelheid klierweefsel uit een direct digitaal verkregen mammogram. Het automatisch vaststellen van deze klierweefseldichtheid is van belang omdat een verhoogde klierweefseldichtheid is erkend als een belangrijke risicofactor voor het ontwikkelen van borstkanker. Verder is een dergelijke klierweefseldichtheidsmaat goed te gebruiken bij het vergelijken van de linker en rechter borst en bij CAD methoden die contextinformatie gebruiken. Er is een fysisch model gebruikt om voor elke pixel in het mammogram de hoeveelheid klierweefsel te berekenen. Dit model gaat ervan uit dat de borst bestaat uit twee typen weefsel, klierweefsel en vet. Door de effectieve lineaire verzwakkingscoefficienten van zowel

klier- als vetweefsel en een borstdikte-correctie toe te passen in het model, is de weefselsamenstelling op een gegeven locatie bepaald. De effectieve lineaire verzwakkingscoefficienten van de twee weefsels types zijn bepaald uit empirische data als functie van de buisspanning, anode en filter materiaal, en de dikte van de gecomprimeerde borst. De methode voor het automatisch bepalen van de klierweefseldichtheid is gevalideerd voor 22 voorbeeld gevallen. Hierbij zijn de klierweefsel volumeschattingen vergeleken met de volumeschattingen afkomstig van semi-automatische segmentatie van klierweefsel verkregen door middel van Magnetic Resonance Imaging (MRI). De correlatie tussen de uit MRI data en de uit mammografie bepaalde volumes was hoog.

Dankwoord

Bij de totstandkoming van dit proefschrift ben ik door vele mensen geholpen. De belangrijkste bijdrage is geleverd door mijn copromotor Nico Karssemeijer. Nico, bedankt voor jouw prettige en kundige begeleiding en het creëren van een prettige werkomgeving gedurende de afgelopen jaren! Daarnaast wil ik mijn promotor Stan Gielen bedanken voor zijn begeleiding en het proeflezen van de hoofdstukken. In het bijzonder wil ik ook Jan Hendriks vermelden die als radioloog betrokken was bij het beschreven onderzoek. Het is heel treurig dat ik hem niet meer persoonlijk kan bedanken. Zijn enthousiasme, kennis op het gebied van de mammografie en betrokkenheid bij het onderzoek waren van zo'n grote invloed dat ze in dit dankwoord zeker genoemd moeten worden.

De vijf jaren die ik heb gewerkt op de afdeling Radiologie van het UMC Nijmegen heb ik zelf als bijzonder plezierig ervaren. Een belangrijke bijdrage hieraan hebben natuurlijk mijn collega's geleverd. Allereerst wil ik hierbij drie directe collega's noemen van wie ik tevens veel hulp heb gekregen bij de totstandkoming van mijn publicaties. Celia, Sheila en Peter bedankt voor al jullie steun en gezelschap! Ton en Sander wil ik graag bedanken voor hun ondersteuning en hulp bij het uitvoeren van experimenten op het screenwerkstation. En ook Henkjan ben ik dank verschuldigd voor onder andere de hulp bij het vergaren van MRI data.

Verder heb ik veel geleerd van het bijwonen van de mammogram - leessessies van Jan Hendriks, Roland Holland, Henny Rijken en Hans Otten. De discussies die dan meestal gevoerd werden over waarom iets nu wel of niet verdacht is in een mammogram zijn ontzettend belangrijk voor het ontwikkelen van CAD technieken. Bedankt dat ik hier bij aanwezig mocht zijn.

Bijzonder was ook de bijdrage van Stichting Bergh in het Zadel. Met een sponsorfietstocht en het uitbrengen van een muziek CD (door de Stichting Barghse Jonges) brachten zij € 340 000 bijeen voor het KWF. Hiermee kon een groot deel van het hier gepresenteerde onderzoek gefinancierd worden.

Tenslotte wil ik iedereen, waaronder mijn vele kamergenoten, die een bijdrage heeft geleverd aan dit proefschrift en de prettige werksfeer bedanken.

Publications

Related to this thesis

Refereed

S. van Engeland, P. Snoeren, J. Hendriks, N. Karssemeijer, *A Comparison of Methods for Mammogram Registration*, IEEE Trans. Med. Imag. 22(11) (2003) 1436-1444.

S. van Engeland, P. Snoeren, H. Huisman, C. Boetes, N. Karssemeijer, *Volumetric Breast Density Estimation from Full Field Digital Mammograms*, IEEE Trans. Med. Imag. 25(3) (2006) 273-282.

S. Timp, S. van Engeland, N. Karssemeijer, *A regional registration method to find corresponding mass lesions in temporal mammogram pairs*, Med. Phys. 32(8) (2005) 2629-2638.

S. van Engeland, S. Timp, N. Karssemeijer, *Finding corresponding regions of interest in mediolateral oblique and craniocaudal mammographic views*, accepted for publication in Med. Phys.

Conference proceedings

S. van Engeland, N. Karssemeijer, *Matching Breast Lesions in Multiple Mammographic Views*, Medical Image Computing and Computer-Assisted Intervention, W. Niessen and M. Viergever, Springer, LNCS, (2001) 1172-1173.

S. van Engeland, N. Karssemeijer, J. Hendriks, *Using information from two mammographic views to improve computer-aided detection of mass lesions*, 6th International Workshop on Digital Mammography, H. Peitgen, Springer-Verlag, Berlin, (2002) 377-381.

S. van Engeland, P. Snoeren, N. Karssemeijer, J. Hendriks, *Optimized Perception of Lesion Growth in Mammograms using Digital Display*, Medical Imaging: Image Perception, Observer Performance, and Technology Assessment, Proceedings of SPIE 5034, D. Chakraborty and E. Krupinski, (2003) 25-31.

S. van Engeland, C. Varela, S. Timp, P. Snoeren, N. Karssemeijer, *Using context for mass detection and classification in mammograms*, Medical Imaging: Image Perception, Observer Performance, and Technology Assessment, Proceedings of SPIE 5749, M.P. Eckstein and Y. Jiang, (2005) 94-102.

S. van Engeland, N. Karssemeijer, *Regrouping initial CAD mass detections to facilitate classification of suspicious regions in mammography*, Medical Imaging: Image Processing, Proceedings of SPIE 5747, J.M. Fitzpatrick and J.M. Reinhardt, (2005) 975-986.

Other

J. Quaedackers, S. van Engeland, P. Mutsaers, J. de Goeij, L. Snoeckx, M. de Voigt, G. van der Vusse, *Determination of the intracellular element concentrations in cultured H9c2 cells by proton microprobe techniques*, Nucl. Instr. and Meth. B 158 (1999) 405-411.

G. Schep, D. Kaandorp, M. Bender, H. Weerdenburg, S. van Engeland, P. Wijn, *Magnetic resonance angiography used to detect kinking in the iliac arteries in endurance athletes with claudication*, Physiol. Meas. 22 (2001) 475-487.

T. Delhaas, S. van Engeland, J. Broers, C. Bouten, N. Kuijpers, F. Ramaekers, L. Snoeckx, *Quantification of cytoskeletal deformation in living cells based on hierarchical feature vector matching*, Am. J. Physiol.; Cell Physiol. 283 (2002) 639-645.

G. Schep, D. Kaandorp, M. Bender, S. van Engeland, H. Weerdenburg, B. Titulaer, P. Wijn, *Excessive length of iliac arteries in athletes with flow limitations measured by magnetic resonance angiography*, Med. Sci. Sports Exer. 34(3) (2002) 385-393.

About the author

Saskia van Engeland was born in Eindhoven, the Netherlands, on November 14, 1976. She completed her secondary education at the Gemeentelijke Scholengemeenschap Woensel in Eindhoven in 1995. In the same year, she started the study Applied Physics at the Eindhoven University of Technology. After she obtained her MSc degree in 2000, she started as junior researcher at the department of Radiology at the University Medical Centre Nijmegen. The work carried out there is presented in this thesis. As of February 1 she is working in the department Software Systems Development of ASML Veldhoven.

Saskia van Engeland werd op 14 november 1976 geboren te Eindhoven. In 1995 behaalde ze haar diploma voor het Voortgezet Wetenschappelijk Onderwijs aan de Gemeentelijke Scholengemeenschap Woensel te Eindhoven. In datzelfde jaar startte zij de studie Technische Natuurkunde aan de Technische Universiteit Eindhoven. Na haar afstuderen in 2000, startte zij als junior onderzoeker bij de afdeling Radiologie van het Universitair Medisch Centrum Nijmegen. Het onderzoek dat ze daar verricht heeft staat beschreven in dit proefschrift. Vanaf 1 februari is zij werkzaam bij de afdeling Software Systems Development van ASML in Veldhoven.